The value of randomized strategies in distributionally robust risk averse network interdiction problems *

Utsav Sadana[1,3] and Erick Delage[†1,2]

[1]GERAD, Montréal, QC H3T 2A7, Canada.
[2]Department of Decision Sciences, HEC Montréal, Montréal, QC H3T 2A7, Canada.
[3]Desautels Faculty of Management, McGill University, Montréal, Québec, H3A 1G5, Canada.

Abstract

Conditional Value at Risk (CVaR) is widely used to account for the preferences of a risk-averse agent in the extreme loss scenarios. To study the effectiveness of randomization in interdiction problems with an interdictor that is both risk and ambiguity averse, we introduce a distributionally robust maximum flow network interdiction problem where the interdictor randomizes over the feasible interdiction plans in order to minimize the worst-case CVaR of the maximum flow with respect to both the unknown distribution of the capacity of the arcs and his own randomized strategy. Using the size of the uncertainty set, we control the degree of conservatism in the model and reformulate the interdictor's distributionally robust optimization problem as a bilinear optimization problem. For solving this problem to any given optimality level, we devise a spatial branch and bound algorithm that uses the McCormick inequalities and reduced reformulation linearization technique to obtain a convex relaxation of the problem. We also develop a column generation algorithm to identify the optimal support of the convex relaxation which is then used in the coordinate descent algorithm to determine the upper bounds. The efficiency and convergence of the spatial branch and bound algorithm is established in the numerical experiments. Further, our numerical experiments show that randomized strategies can have significantly better performance than optimal deterministic ones.

1 Introduction

In a deterministic maximum flow network interdiction problem (MFNIP), the interdictor, with a fixed budget of interdiction, removes arcs from a capacitated network with the aim to minimize the maximum flow that can be routed by the adversary (flow player) in the network. In Wood (1993), the author showed that MFNIP is NP-hard. In many real-world applications, the parameters of an agent's decision model, e.g., the capacity of the arcs in a network interdiction problem, can be uncertain. It was shown in Ben-Tal and Nemirovski (2000) that perturbations in the parameters of a linear programming problem can render the solution to become infeasible or significantly suboptimal. A popular technique to tackle this issue is Stochastic

*This article was reviewed and accepted in INFORMS Journal on Computinghttps://doi.org/10.1287/ijoc.2022.1257 .

[†]Email addresses: *utsav.sadana@mail.mcgill.ca*, *erick.delage@hec.ca*

Programming (SP) which assumes that the distribution of the parameters is known. However, in many instances, the distribution of the random parameters is not exactly known. This can lead to post-decision disappointment referred to as the *optimizer's curse* (see, Smith and Winkler (2006)). Alternatively, classical robust optimization models do not use any distributional information on the uncertain parameters. However, imposing that only the worst-case outcome is to be taken into account can often lead to over-conservative solutions.

Distributionally Robust Optimization (DRO) acts as a compromise between SP and RO by requiring the probability distribution to lie in a distributional ambiguity set and seeking a solution that performs best according to the worst-case distribution. Hence, when the ambiguous distribution set is properly tuned, it can prevent both the post-decision disappointment of SP models and the over-conservatism of RO models. Furthermore, the set can be calibrated in ways that will provide statistical guarantees on the out-of-sample performance of the DRO solution. Recently, it was shown in Delage et al. (2019) that it can be beneficial to employ randomization in non-convex DRO problems. For an ambiguity-averse risk-neutral decision maker, i.e. one that minimizes worst-case expected value, Delage and Saif (2021) proposed an algorithm to identify such strategies in mixed-integer two-stage DRO problems. Bertsimas et al. (2016) also studied the value of randomization specifically in the context of a network interdiction game with known parameters and risk-neutral agents. Nevertheless, none of these works provide either theoretical or computational means of identifying optimal randomized solutions for agents that employ more general risk measures than expected value, such as the Conditional Value at Risk (CVaR) measure introduced in Rockafellar and Uryasev (2000). Such risk measures are especially relevant in security policy models such as network interdiction problems where a decision maker, e.g., a law-enforcement agency, might be concerned by the possibility of incurring huge losses under certain scenarios, e.g. caused by a large flow of illegal drugs, weapons or money.

In this paper, we study a *distributionally robust maximum flow network interdiction problem* where the interdictor employs a CVaR risk measure to model his risk aversion. Our contributions can be summarized as follows:

- On the methodological side, we introduce for the first time ambiguity and risk aversion in network interdiction games where the interdictor minimizes the worst-case CVaR over both the unknown distribution of the capacities of the arcs and the distribution of interdicted arcs. This is in sharp contrast with the work of Loizou (2015) who considers an interdictor that employs CVaR to handle parameter uncertainty but an expected value to handle the uncertainty caused by his own randomized strategy. We show that the approach in Loizou (2015) can in fact produce a solution that is stochastically strictly dominated by the solution of our proposed model.
- On the algorithmic side, we complement the work of Delage and Saif (2021) by designing the first algorithm that can identify optimal randomized strategies for an ambiguity and risk averse agent, i.e. an agent that minimizes a worst-case convex risk measure other than the expected value. Our algorithm is based on a spatial branch and bound scheme (see, Al-Khayyal and Falk (1983)) embedded with a novel implementation of the column generation (CG) method designed to exploit the structure of a linearized large scale bilinear optimization problem. It will successfully identify high quality randomized solution for networks containing hundreds of nodes in a few minutes.
- On the empirical side, we provide evidence which indicates that a network interdictor can significantly benefit from randomization. We find that, while it is true that deterministic plans are often worst-case CVaR optimal, in instances where randomization strictly improves this objective, the improvement is significant, i.e. up to 12% average in-sample improvement depending on conditions. Furthermore, we also find evidence that for instances where a randomized strategy significantly improves the in-sample performance, there is benefit in employing randomization on the out-of-sample scenarios.

The rest of the paper is organized as follows. Section 2 gives an overview of the literature covering network interdiction and DRO problems which are closely related to this work. We also briefly discuss the algorithms which have been proposed previously to solve non-convex optimization problems with bilinear constraints. Section 3 defines our *distributionally robust maximum flow network interdiction problem*. The robust counterpart of the DRO model is given in Section 4 where we also describe the spatial branch and bound algorithm embedded with the CG algorithm that is used to solve the interdictor's problem to global optimality. Numerical experiments are provided in Section 5 to demonstrate the convergence and efficiency of our algorithm, and provide evidence that randomization can significantly improve the performance achieved by deterministic interdiction plans. Concluding remarks are given in Section 6. Finally, all proofs can be found in the appendix.

Remark 1. *We take a pause to summarize Remark 1 from Delage and Saif (2021), which clarifies the role of "risk aversion" and "ambiguity aversion" in decision making under uncertainty with randomized strategies. Namely, as formalized in Ellsberg (1961) and Epstein (1999), risk aversion refers to how a decision maker reacts "to situations where the perceived likelihood of events of interests can be represented by probabilities", whereas ambiguity aversion refers to her way of handling "situations where the information available is too imprecise to be summarized by a probability measure". Both types of aversion need to be accounted for in decision problems with distributional ambiguity when a randomized strategy is used given that the latter can generate random variables that either have known or unknown distributions. In this work, we will employ a CVaR measure to characterize risk aversion of the random maximum flows with known distributions, while we employ the axiomatic motivation from Delage et al. (2019) to justify the use of worst-case CVaR when the distribution of the random maximum flow is ambiguous.*

Notations

Vectors are expressed in bold and matrices are represented by capital letters. $\mathbf{1}$ and $\mathbf{0}$ denote column vectors of 1's and 0's respectively. The identity matrix is denoted by $\mathbb{I}$, and $\boldsymbol{e}_i$ captures its i-th column. The set of all probability measures on a finite discrete measurable space $(\mathcal{X}, F_{\mathcal{X}})$ is denoted by $\Delta\mathcal{X} \subseteq \mathbb{R}_+^{|\mathcal{X}|}$, where $F_{\mathcal{X}}$ denotes all subsets of $\mathcal{X}$.

2 Related Literature

It is well-known that the illegal flow of drugs, weapons or other hazardous substances poses a threat to the security of a nation, see, Magliocca et al. (2019) and the references therein. The law-enforcement agencies aim to reduce their flow while the adversaries, which may be smugglers or terrorist organizations, try to increase it. Network interdiction problems find useful applications in transportation (Israeli and Wood (2002)), medicine (Assimakopoulos (1987)), and disruption of supply chains of illicit drugs (McLay et al. (2011)). The interdiction problems with a defender who interdicts a set of arcs on the network to minimize the flow while the adversary maximizes it are known as maximum flow network interdiction problems (see Wood (1993), Cormican et al. (1998), Smith et al. (2013), Smith and Song (2020)). The maximum flow network interdiction problems are one of the most widely studied network flow problems. For an elaborate description of the theory, algorithms and applications of network flow problems, refer to Ahuja et al. (1993).

Recently, a series of papers have considered risk aversion in network interdiction problems. Lei et al. (2018) studied a maximum flow network interdiction problem where interdictor and follower use deterministic strategies and the effect of interdiction on the capacity of each arc is random. Also, the interdictor and the follower use the CVaR risk measure. In Atamtürk et al. (2020), the capacities are assumed to be stochastic, and the interdictor minimizes the maximum flow-at-risk over a discrete set of actions. We note that there is another stream of literature on

shortest-path interdiction problems where the leader removes arcs to increase the length of the shortest path in the network (Israeli and Wood 2002). In Song and Shen (2016), the authors study a risk averse shortest path network interdiction problem with chance constraint where the risk averse interdictor aims to ensure with high probability that the length of shortest path is above a threshold. Chicoisne et al. (2018) provide decomposition-based methods to solve shortest path problems with CVaR and entropic risk measure. In Pay et al. (2019), the defender is ambiguous about her own utility function and determines her decision by optimizing over the worst utility function from a set of utility functions constructed from the historical data. In Borrero et al. (2019) and Yang et al. (2021), sequential shortest-path network interdiction problems are studied where the interdictor learns the cost of the edges by observing the actions of the follower over time. On the other hand, we consider a one-shot maximum flow network interdiction problem in this paper.

Another topic that has received attention in the network interdiction literature is on determining the value of randomization. In Jain et al. (2010), it is shown that randomization can be useful in security applications like patrolling of airports. Bertsimas et al. (2016) introduced a randomized network interdiction game where only the interdictor can randomize and where the follower observes the distribution of interdicted arcs. The authors assumed that the interdictor as well as the flow player are risk-neutral, and the model parameters are known with certainty. Our setting can be considered as a variation of their work where the value of randomization stems from the leader's ambiguity aversion rather than from the assumption that the follower is partially informed about the interdictor's action. We also note that Holzmann and Smith (2021) studied the value of randomization in a shortest-path interdiction problem for a risk-neutral leader. The authors neither consider ambiguity nor risk aversion.

Table 1: Maximum flow interdiction models.

Reference	Risk measure for interdictor	Interdiction Strategies	Capacities
Cormican et al. (1998)	Expectation	Deterministic	Stochastic
Janjarassuk and Linderoth (2008)	Expectation	Deterministic	Stochastic
Bertsimas et al. (2016)	Expectation	Randomized	Known
Lei et al. (2018)	CVaR	Deterministic	Known
Atamtürk et al. (2020)	Value at Risk (VaR)	Deterministic	Stochastic (Normal distribution)
Our model	CVaR	Randomized	Unknown (Ambiguous distribution)

Clearly, in the one-shot interdiction problems, the probability distribution over the uncertain quantities is either assumed to be known or a worst-case approach is used to determine optimal strategies (see Smith and Song (2020) for a detailed exposition of the interdiction models). Also, the benefit of using randomized strategies for a risk-averse agent in DRO problems has not been explored as can be seen in Table 1 . In order to account for ambiguity aversion in non-cooperative games, Loizou (2015) has proposed a distributionally robust game theory model where players use a worst-case CVaR of expected payoff to evaluate the performance of their mixed strategies. We show, using an example, that a single player counterpart of his model results in an optimal policy that is strictly stochastically dominated by the strategy produced by our model.

Alternatively, we propose the *distributionally robust maximum flow network interdiction model* where the interdictor minimizes the worst-case CVaR of the maximum flow with respect to both the unknown distribution of the capacity of the arcs and his randomized strategy over the feasible interdiction plans. Similar to Janjarassuk and Linderoth (2008), we assume that the flow player maximizes the flow under any interdiction plan after observing the capacity of each arc. However, the success of interdiction is a Bernoulli random variable in Janjarassuk and

Linderoth (2008) while it is deterministic in our model.

Hajinezhad and Shi (2018) and Hajinezhad and Hong (2019) solved non-convex non-smooth optimization problems with bilinear equality constraints for machine learning and signal processing applications using the alternating direction method of multipliers. Their algorithm exploits the separability of the objective function in the decision variables to obtain local optimal solutions efficiently. On the other hand, the decision variables in our problem are coupled, and therefore, we devise the spatial branch and bound algorithm (see Al-Khayyal and Falk (1983)) that iteratively searches the feasible space of the problem to obtain global optimal solutions within a given tolerance. The rate of convergence of the algorithm relies on generating tighter lower and upper bounds. In the literature, McCormick inequalities (see, McCormick (1976)) are commonly used to obtain convex relaxations of the non-convex programming problems with bilinear constraints. In order to obtain a tighter bound, Reformulation Linearization Technique (RLT) was proposed in Sherali and Alameddine (1992) wherein valid inequalities obtained by multiplying the pairs of feasible constraints are added to the relaxed problem. This increases the size of the LP relaxation. Liberti and Pantelides (2006) proposed the reduced RLT, which in contrast to RLT, can give an exact reformulation of original bilinear problem with an additional number of linear equality constraints. The authors also show that the reduced RLT combined with McCormick inequalities can result in tighter relaxations than applying McCormick directly to the bilinear terms.

3 Distributionally robust maximum flow network interdiction problem

Consider a directed graph $G = (V, E)$, where V and E denote the nodes and arcs, respectively. Let $e = (i, j)$ represent an arc on G, $\delta^-(i) = \{(i, j)|j \in V\}$ and $\delta^+(i) = \{(j, i)|j \in V\}$ denote, respectively, the set of arcs leaving and entering node $i \in V$. A flow in the graph G is denoted by a non-negative vector $\boldsymbol{x} \in \mathbb{R}_+^{|E|}$ so that for each $e \in E$, we have $x_e \leq c_e$, where the capacity of all arcs is denoted by $\boldsymbol{c} \in \mathbb{R}_+^{|E|}$. The conservation of flow at each node is ensured by

$$\sum_{e \in \delta^-(i)} x_e - \sum_{e \in \delta^+(i)} x_e = 0 \qquad \forall i \notin \{\mathfrak{s}, \mathfrak{t}\}, \tag{1}$$

where $\mathfrak{s}$ and $\mathfrak{t}$ denote the source and sink node, respectively. The flow player aims at maximizing the flow in the network. The interdictor, on the other hand, aims at minimizing the worst-case CVaR of the flow with respect to both the unknown distribution of the capacity of the arcs and his mixed strategy over the interdicted arcs. We assume that the interdictor has a budget to remove B arcs in the network. Let $\mathcal{L}$ denote the finite set of feasible plans for the interdictor where $\mathcal{L} = \{\boldsymbol{\ell} \in \{0, 1\}^{|E|} | \sum_e \ell_e \leq B\}$. Here, $\ell_e = 1$ if the interdictor removes arc e and $\ell_e = 0$ if arc e is not interdicted. The distribution of the capacities of all the arcs is only known to lie in the set $\mathcal{Q}$. We assume that the distribution is discrete with a set of scenarios $\mathcal{K}$ supported on $\{\boldsymbol{c}^k\}_{k \in \mathcal{K}}$.

Similar to the model in Bertsimas et al. (2016), we assume that the randomized strategy of the interdictor is a probability distribution $\boldsymbol{u}$ over the set $\mathcal{L}$ where $\boldsymbol{u} \in \Delta\mathcal{L}$. For any interdiction plan $\boldsymbol{\ell}$ and scenario k, the flow player solves the following problem

$$f_{\boldsymbol{\ell},k} := \underset{\boldsymbol{x} \in \mathbb{R}^{|E|}}{\text{maximize}} \quad \boldsymbol{d}^\top \boldsymbol{x} \tag{2a}$$

$$\text{subject to} \quad N\boldsymbol{x} = 0 \tag{2b}$$

$$0 \leq \boldsymbol{x} \leq C^k(\mathbf{1} - \boldsymbol{\ell}), \tag{2c}$$

where $\boldsymbol{d}^\top \boldsymbol{x} = \sum_{e \in \delta^+(t)} x_e$, $C^k = \text{diag}(\boldsymbol{c}^k)$ and N is the node-arc incidence matrix of the network without the rows associated to nodes $\mathfrak{s}$ and $\mathfrak{t}$.

The interdictor solves the following distributionally robust maximum flow network interdiction (DRMFNI) problem:

$$\text{(DRMFNI)} \quad \underset{\boldsymbol{u}\in\Delta\mathcal{L}}{\text{minimize}} \max_{\boldsymbol{q}\in\mathcal{Q}} \text{CVaR}^{\alpha}_{\boldsymbol{\ell}\sim\boldsymbol{u},k\sim\boldsymbol{q}}[f_{\boldsymbol{\ell},k}], \tag{3}$$

where CVaR is defined over the joint distribution of capacities and interdicted arcs. Namely, with risk aversion parameter $\alpha = [0, 1)$, CVaR is defined as

$$\text{CVaR}^{\alpha}_{\boldsymbol{\ell}\sim\boldsymbol{u},k\sim\boldsymbol{q}}[f_{\boldsymbol{\ell},k}] := \inf_{\zeta} \ \zeta + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k u_{\boldsymbol{\ell}}[f_{\boldsymbol{\ell},k} - \zeta]^+,$$

where $[f_{\boldsymbol{\ell},k} - \zeta]^+ := \max(f_{\boldsymbol{\ell},k} - \zeta, 0)$ (see Rockafellar and Uryasev (2000)).

Remark 2. *Since the set* $\mathcal{X} := \{\boldsymbol{x}\in\mathbb{R}^{|E|} \mid N\boldsymbol{x} = 0,\ 0 \le \boldsymbol{x} \le C^k(\mathbf{1} - \boldsymbol{\ell})\}$ *of all possible* $\mathfrak{s}-\mathfrak{t}$ *flows is convex, randomization is not beneficial for the flow player if he considers minimizing a convex risk measure of* $\boldsymbol{d}^\top\boldsymbol{x}$*, such as,* $\min_{\boldsymbol{u}_x\in\Delta\mathcal{X}} CVaR^{\alpha}_{\boldsymbol{x}\sim\boldsymbol{u}_x}[-\boldsymbol{d}^\top\boldsymbol{x}]$ *(see, Delage et al. (2019)).*

Under this setting, one can show that the interdictor's DRMFNI problem can be cast as a bilinear DRO problem.

Proposition 1. *When* $\mathcal{Q}$ *is a convex set, the interdictor's DRMFNI problem* (3) *is equivalent to the following bilinear DRO problem*

$$\begin{aligned}
\underset{\boldsymbol{u},\zeta,\Delta,t,\boldsymbol{\eta}}{\text{minimize}} \quad & t && && \text{(4a)}\\
\textit{subject to} \quad & \zeta + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k\Delta_{\boldsymbol{\ell},k} \le t && \forall \boldsymbol{q}\in\mathcal{Q} && \text{(4b)}\\
& \Delta_{\boldsymbol{\ell},k} \ge u_{\boldsymbol{\ell}} f_{\boldsymbol{\ell},k} - \eta_{\boldsymbol{\ell}} && \forall \boldsymbol{\ell}\in\mathcal{L},\ k\in\mathcal{K} && \text{(4c)}\\
& \eta_{\boldsymbol{\ell}} = u_{\boldsymbol{\ell}}\zeta && \forall \boldsymbol{\ell}\in\mathcal{L} && \text{(4d)}\\
& \Delta_{\boldsymbol{\ell},k} \ge 0 && \forall \boldsymbol{\ell}\in\mathcal{L},\ k\in\mathcal{K} && \text{(4e)}\\
& \boldsymbol{u} \ge 0 && && \text{(4f)}\\
& \mathbf{1}^\top\boldsymbol{u} = 1 && && \text{(4g)}\\
& 0 \le \zeta \le \bar{\zeta}, && && \text{(4h)}
\end{aligned}$$

where $\bar{\zeta} := \max_{k\in\mathcal{K}} f_{\mathbf{0},k}$.

The above problem is non-convex due to the bilinear terms $u_{\boldsymbol{\ell}}\zeta$. A second difficulty resides in having to compute $f_{\boldsymbol{\ell},k}$ for each scenario $k \in \mathcal{K}$ and interdiction plan $\boldsymbol{\ell} \in \mathcal{L}$ in order to solve (4). This will be addressed algorithmically in Section 4.

In the following example, we show that the distributionally robust model proposed in Loizou (2015) identifies interdiction strategies which are strictly stochastically dominated by other feasible strategies. This indicates that Loizou's approach is not well motivated for this class of problems. Recall that Loizou (2015) considers an interdictor that employs CVaR to handle parameter uncertainty but an expected value to handle the uncertainty caused by his own randomized strategy.

Example 1. *Consider an agent trying to reduce the maximum flow from point* $\mathfrak{s}$ *to point* $\mathfrak{t}$ *which are located on two respective sides of a river. The agent has a budget to interdict two routes. There are three routes available,* $e \in \{T1,\ T2,\ B\}$*, where routes* $T1$ *and* $T2$ *use two different tunnels to pass the river, while route* B *uses a bridge to do so. In normal traffic conditions, the capacities are,* τ*,* τ*, and* ϵ*, with* $\frac{2\tau}{3} < \epsilon < \tau$*, for routes* $T1$*,* $T2$*, and* B *respectively. Unfortunately, all three routes are susceptible to congestion on the day of interest. In the case of* $T1$ *and* $T2$*, it is known that the city is planning to do some repairs, which would decrease the*

flow by δ, using one repair team but no information is available regarding which tunnel it will be; the identity of the selected tunnel is denoted by $r \in \{1,\ 2\}$. On the other hand, there is also a weather forecast that predicts 50% chance of snowfall which would create the same decrease in flow, δ, on the bridge used by route B. We let δ satisfy $2(\tau-\epsilon) < \delta \leq \epsilon$. The flow player wants to maximize the flow from point $\mathfrak{s}$ to point $\mathfrak{t}$.

Let the set of feasible interdiction plans be defined as follows:

$$\mathcal{L} = \{\{T1, B\}, \{T2, B\}, \{T1, T2\}, \{T1\}, \{T2\}, \{B\}, \{\varnothing\}\}.$$

The possible scenarios $k = \{1, 2, 3, 4\}$ for the capacity of the three routes $T1$, $T2$ and B are given by:

$$\boldsymbol{c}^1 = \begin{bmatrix} \tau-\delta \\ \tau \\ \epsilon-\delta \end{bmatrix},\ \boldsymbol{c}^2 = \begin{bmatrix} \tau \\ \tau-\delta \\ \epsilon-\delta \end{bmatrix},\ \boldsymbol{c}^3 = \begin{bmatrix} \tau-\delta \\ \tau \\ \epsilon \end{bmatrix},\ \boldsymbol{c}^4 = \begin{bmatrix} \tau \\ \tau-\delta \\ \epsilon \end{bmatrix},$$

with respective probabilities q_1, q_2, q_3, q_4, such that $q_1 + q_2 = 0.5$ and $q_3 + q_4 = 0.5$.

Here are the numerical details regarding $f(\boldsymbol{\ell}, k)$ which denotes the total flow when interdiction plan $\boldsymbol{\ell}$ is chosen, and scenario k is realized:

$$f(\boldsymbol{\ell}, k) := \begin{cases} \tau & \text{if } \boldsymbol{\ell} = \{T1, B\} \text{ and } k \in \{1, 3\}, \\ \tau-\delta & \text{if } \boldsymbol{\ell} = \{T1, B\} \text{ and } k \in \{2, 4\}, \\ \tau-\delta & \text{if } \boldsymbol{\ell} = \{T2, B\} \text{ and } k \in \{1, 3\}, \\ \tau & \text{if } \boldsymbol{\ell} = \{T2, B\} \text{ and } k \in \{2, 4\}, \\ \epsilon-\delta & \text{if } \boldsymbol{\ell} = \{T1, T2\} \text{ and } k \in \{1, 2\}, \\ \epsilon & \text{if } \boldsymbol{\ell} = \{T1, T2\} \text{ and } k \in \{3, 4\}, \\ \tau+\epsilon-\delta & \text{if } \boldsymbol{\ell} = \{T1\} \text{ and } k \in \{1, 4\}, \\ \tau+\epsilon-2\delta & \text{if } \boldsymbol{\ell} = \{T1\} \text{ and } k = 2, \\ \tau+\epsilon & \text{if } \boldsymbol{\ell} = \{T1\} \text{ and } k = 3, \\ \tau+\epsilon-2\delta & \text{if } \boldsymbol{\ell} = \{T2\} \text{ and } k = 1, \\ \tau+\epsilon-\delta & \text{if } \boldsymbol{\ell} = \{T2\} \text{ and } k \in \{2, 3\}, \\ \tau+\epsilon & \text{if } \boldsymbol{\ell} = \{T2\} \text{ and } k = 4, \\ 2\tau-\delta & \text{if } \boldsymbol{\ell} = \{B\} \text{ and } k \in \{1, 2, 3, 4\}, \\ 2(\tau-\delta)+\epsilon & \text{if } \boldsymbol{\ell} = \{\varnothing\} \text{ and } k \in \{1, 2\}, \\ 2\tau-\delta+\epsilon & \text{if } \boldsymbol{\ell} = \{\varnothing\} \text{ and } k \in \{3, 4\}. \end{cases}.$$

Consider two potential strategies:

$$\boldsymbol{u}^L = (0.5, 0.5, 0, 0, 0, 0, 0), \qquad \boldsymbol{u}^{SD} = (0, 0, 1, 0, 0, 0, 0),$$

where $\boldsymbol{u}^L$ denotes the strategy to interdict routes $(T1, B)$ with 50% probability and routes $(T2, B)$ with 50% probability; $\boldsymbol{u}^{SD}$ denotes the strategy to interdict routes $(T1, T2)$ with probability 1.

Our robust risk-averse approach $g^{SD}(\boldsymbol{u}) := \sup_{\boldsymbol{q}\in\mathcal{Q}} CVaR^{\alpha}_{\boldsymbol{\ell}\sim\boldsymbol{u}, k\sim\boldsymbol{q}}[f(\boldsymbol{\ell}, k)]$, with $\alpha \geq 50\%$ leads to the following evaluation:

$$g^{SD}(\boldsymbol{u}^L) = \tau \qquad \textit{v.s.} \qquad g^{SD}(\boldsymbol{u}^{SD}) = \epsilon.$$

Since $\epsilon < \tau$, we get that $\boldsymbol{u}^{SD}$, interdicting both routes $T1$ and $T2$, outperforms $\boldsymbol{u}^L$.

Alternatively, the approach in Loizou (2015) can be summarized as minimizing $g^L(\boldsymbol{u}) := \sup_{\boldsymbol{q}\in\mathcal{Q}} CVaR^{\alpha}_{k\sim\boldsymbol{q}}[\mathbb{E}_{\boldsymbol{\ell}\sim\boldsymbol{u}}[f(\boldsymbol{\ell}, k)]]$ and leads to the following evaluation

$$g^L(\boldsymbol{u}^L) = \tau - \delta/2 \qquad \textit{v.s.} \qquad g^L(\boldsymbol{u}^{SD}) = \epsilon.$$

Since $\delta > 2(\tau-\epsilon)$, it is strictly better to implement $\boldsymbol{u}^L$, i.e. interdicting $\{T1, B\}$ with 50% probability and $\{T2, B\}$ with 50% probability, outperforms $\boldsymbol{u}^{SD}$. Yet, it is clear from a purely

statistical point of view that $\boldsymbol{u}^{SD}$ *strictly stochastically dominates* $\boldsymbol{u}^{L}$ *for all* $\boldsymbol{q} \in \mathcal{Q}$*. Specifically, we have that*

$$\begin{aligned}\forall t \in \mathbb{R},\ \mathbb{P}_{\boldsymbol{\ell}\sim\boldsymbol{u}^{L},k\sim\boldsymbol{q}}(f(\boldsymbol{\ell},k) \geq t) &= 0.5\mathbb{1}_{t\leq\tau} + 0.5\mathbb{1}_{t\leq\tau-\delta} \\ &\geq 0.5\mathbb{1}_{t\leq\epsilon} + 0.5\mathbb{1}_{t\leq\epsilon-\delta} = \mathbb{P}_{\boldsymbol{\ell}\sim\boldsymbol{u}^{SD},k\sim\boldsymbol{q}}(f(\boldsymbol{\ell},k) \geq t)\,,\end{aligned}$$

where $\mathbb{1}_x$ *denotes the indicator function of set* x*, and where inequality is strict when* $\epsilon < t \leq \tau$ *or* $\epsilon - \delta < t \leq \tau - \delta$*.*

4 Solving the DRMFNI problem

In this section, we propose a numerical scheme for solving the DRMFNI problem. It is well-known that the tractability of a robust optimization problem depends on the structure of the uncertainty set (see Ben-Tal and Nemirovski (2008) and Ben-Tal et al. (2015)). In particular, for simplicity of exposition, we will focus without loss of generality on the case where the distribution ambiguity set contains perturbed versions of a reference distribution $\hat{\boldsymbol{q}} \in \Delta\mathcal{K}$.

Assumption 1. *Let* $\mathcal{Q}$ *be defined as follows:*

$$\mathcal{Q} := \left\{ \boldsymbol{q} \in \mathbb{R}^{|\mathcal{K}|} \ | \exists \boldsymbol{z} \in \mathcal{Z},\ \ \boldsymbol{q} \geq 0, \sum_{k=1}^{|\mathcal{K}|} q_k = 1, \boldsymbol{q} = \hat{\boldsymbol{q}} + \boldsymbol{z} \right\},$$

where $\hat{\boldsymbol{q}} \in \Delta\mathcal{K}$ *is a reference distribution,* $\boldsymbol{z}$ *denotes the perturbation in the reference distribution, and* $\mathcal{Z}$ *is a convex and compact set with* $\mathbf{0}$ *in the relative interior of* $\mathcal{Z}$*.*

The choice of set $\mathcal{Q}$ takes into account the ambiguity of the interdictor regarding the distribution over the capacity of arcs as it consists of probability distributions that are perturbations of the reference distribution.

To convert the DRMFNI problem given in (4) to a finite dimensional optimization problem, we first derive the Fenchel robust counterpart (Ben-Tal et al. (2015)) of the linear constraint (4b) as presented in the following proposition.

Proposition 2. *Let Assumption 1 hold. The robust counterpart of the interdictor's DRMFNI problem presented in* (4) *is given by*

$$\underset{\boldsymbol{u},t,\boldsymbol{\gamma},\Delta,\zeta,\boldsymbol{\eta}}{\text{minimize}} \quad t \tag{5a}$$

$$\text{subject to} \quad \hat{\boldsymbol{q}}^{\top}\boldsymbol{\gamma} + \delta^{*}(\boldsymbol{\gamma}|\mathcal{Z}) - \gamma_k + \zeta + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}\in\mathcal{L}} \Delta_{\boldsymbol{\ell},k} \leq t \qquad \forall k \in \mathcal{K}\,, \tag{5b}$$

$$(4c) - (4h),$$

where $\delta^{*}(\boldsymbol{\gamma}|\mathcal{Z}) := \sup_{\boldsymbol{z}\in\mathcal{Z}} \boldsymbol{\gamma}^{\top}\boldsymbol{z}$ *denotes the support function of the set* $\mathcal{Z}$*.*

The support function of $\mathcal{Z}$ can be obtained for important choices of uncertainty sets $\mathcal{Z}$, e.g., polyhedral, ellipsoidal and second order cone representable sets, see Ben-Tal et al. (2015).

In the rest of the paper, we assume that the set of perturbations $\mathcal{Z}$ refers to the following set:

$$\mathcal{Z}(\Gamma) = \left\{ \boldsymbol{z} \in \mathbb{R}^{|\mathcal{K}|} \,\middle|\, \|\boldsymbol{z}\|_{\infty} \leq \Gamma, \|\boldsymbol{z}\|_1 \leq \Gamma\sqrt{|\mathcal{K}|} \right\}, \tag{6}$$

where $\mathcal{Z}(\Gamma)$ is chosen for simplicity of presentation as a polyhedral set conveniently used to approximate a L2-norm ball of radius Γ. By changing the parameter Γ, the degree of conservatism in the model can be controlled which enables the interdictor to evaluate the trade-off

between robustness and performance. A valuable property of using this uncertainty set is that the robust counterpart of a linear constraint is representable in a linear program. This implies that the bilinear DRO problem (4) can be reformulated as a finite-dimensional bilinear optimization problem. A similar reformulation also necessarily exists for other polyhedral sets such as some of the forms that are based on hypothesis testing (see e.g. Kolmogorov-Smirnov and Wasserstein ambiguity sets in Postek et al. (2016)).[1]

Proposition 3. *When the set of perturbations $\mathcal{Z}$ is given by* (6)*, the robust counterpart of the interdictor's DRMFNI problem given in* (4) *is a bilinear optimization problem*

$$\begin{aligned}
\underset{\substack{\boldsymbol{u},t,\boldsymbol{w}^+,\boldsymbol{w}^-,\beta\\ \boldsymbol{\eta},\boldsymbol{\theta}^+,\boldsymbol{\theta}^-,\Delta,\zeta}}{\text{minimize}} \quad & t && && \text{(7a)}\\
\text{subject to} \quad & \zeta + \Gamma\sqrt{|\mathcal{K}|}\beta + \Gamma\sum_{k'\in\mathcal{K}}(w^+_{k'}+w^-_{k'}) + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}\in\mathcal{L}}\Delta_{\boldsymbol{\ell},k} && && \\
& \quad +(\hat{\boldsymbol{q}}-\boldsymbol{e}_k)^\top(\boldsymbol{w}^+-\boldsymbol{w}^-+\boldsymbol{\theta}^+-\boldsymbol{\theta}^-) \le t && \forall k\in\mathcal{K} && \text{(7b)}\\
& \theta^+_k+\theta^-_k=\beta && \forall k\in\mathcal{K} && \text{(7c)}\\
& \boldsymbol{w}^+\ge 0,\ \boldsymbol{w}^-\ge 0,\ \boldsymbol{\theta}^+\ge 0,\ \boldsymbol{\theta}^-\ge 0 && && \text{(7d)}\\
& \text{(4c)}-\text{(4h)}. && &&
\end{aligned}$$

The procedure that we propose for solving problem (7) is motivated by the observation that ζ and $\boldsymbol{u}$ are complicating variables. Indeed, when either ζ or $\boldsymbol{u}$ is fixed, the problem reduces to a linear program. Since ζ is also constrained to lie in a bounded interval $\bar{\mathcal{I}} := [0,\ \bar{\zeta}]$, a spatial branch and bound scheme on ζ seems appropriate (see, Chandraker and Kriegman (2008)).

Our implementation of the spatial branch and bound algorithm will rely on the existence of two operators. Namely, after defining the problem that we are interested in solving as

$$\begin{aligned}
g(\bar{\mathcal{I}}) := \quad & \min_{\substack{\boldsymbol{u},t,\boldsymbol{w}^+,\boldsymbol{w}^-,\beta,\\ \boldsymbol{\eta},\boldsymbol{\theta}^+,\boldsymbol{\theta}^-,\Delta,\zeta}} && t\\
& \text{subject to} && \text{(4c)}-\text{(4g)},\text{(7b)}-\text{(7d)},\\
& && \zeta\in\bar{\mathcal{I}}\,,
\end{aligned}$$

the algorithm will assume the existence of the following two bounding operators $g_{\text{lb}}(\mathcal{I})$ and $g_{\text{ub}}(\mathcal{I})$ which satisfy:

$$g_{\text{lb}}(\mathcal{I}) \le g(\mathcal{I}) \le g_{\text{ub}}(\mathcal{I})\,,\ \forall \mathcal{I}\subseteq\bar{\mathcal{I}}\,,$$

and such that for all sequence of intervals $\mathcal{I}_1,\mathcal{I}_2,\dots$ converging to some ζ, the associated sequence of bounds $(g_{\text{lb}}(\mathcal{I}_j), g_{\text{ub}}(\mathcal{I}_j))$ converge to $g(\zeta)$. Finally, the operator $g_{\text{ub}}(\mathcal{I})$ will be such that one can always efficiently produce a $\zeta^*_{\text{ub}}(\mathcal{I})\in\mathcal{I}$ such that $g(\{\zeta^*_{\text{ub}}(\mathcal{I})\}) = g_{\text{ub}}(\mathcal{I})$.

With this in hand, we can describe the algorithm. First in words, the spatial branch and bound algorithm starts at a root node capturing $\bar{\mathcal{I}}$ and branches on this node by subdividing it into a number of sub-intervals, considered sub-nodes of the branch and bound tree[2]. Nodes are

[1] In fact, (6) can be seen as a polyhedral approximation of an ambiguity set based on Pearson's χ^2 test when $\hat{\boldsymbol{q}}$ is uniform.

[2] Implementation details: for any interval $\tilde{\mathcal{I}} := [\zeta_{\text{lb}},\ \zeta_{\text{ub}}]$, we create 3 sub-intervals $[\zeta_{\text{lb}},\ \zeta_{\text{lb}}+p_1(\zeta^*_{\text{ub}}(\tilde{\mathcal{I}}))-\zeta_{\text{lb}})]$, $[\zeta_{\text{lb}}+p_1(\zeta^*_{\text{ub}}(\tilde{\mathcal{I}}))-\zeta_{\text{lb}}),\ \zeta^*_{\text{ub}}(\tilde{\mathcal{I}}))+p_2(\zeta_{\text{ub}}-\zeta^*_{\text{ub}}(\tilde{\mathcal{I}})))]$ and $[\zeta^*_{\text{ub}}(\tilde{\mathcal{I}}))+p_2(\zeta_{\text{ub}}-\zeta^*_{\text{ub}}(\tilde{\mathcal{I}}))),\ \zeta_{\text{ub}}]$, where $p_1=\frac{8}{9}$ and $p_2=\frac{1}{9}$ if $\zeta_{\text{lb}}<\zeta^*_{\text{ub}}(\tilde{\mathcal{I}}))<\zeta_{\text{ub}}$; we create subintervals $[\zeta_{\text{lb}},\ \zeta_{\text{lb}}+p_1(\zeta_{\text{ub}}-\zeta_{\text{lb}})]$, $[\zeta_{\text{lb}}+p_1(\zeta_{\text{ub}}-\zeta_{\text{lb}}),\ \zeta_{\text{lb}}+p_2(\zeta_{\text{ub}}-\zeta_{\text{lb}})]$ and $[\zeta_{\text{lb}}+p_2(\zeta_{\text{ub}}-\zeta_{\text{lb}}),\ \zeta_{\text{ub}}]$ where $p_1=\frac{1}{9}$ and $p_2=\frac{5}{9}$ if $\zeta_{\text{lb}}=\zeta^*_{\text{ub}}(\tilde{\mathcal{I}}))$, otherwise $p_1=\frac{4}{9}$ and $p_2=\frac{8}{9}$. This scheme creates partitions such that after branching, the smallest partition contains optimal $\zeta^*_{\text{ub}}(\tilde{\mathcal{I}})$ for the upper bounding problem in the interval $\tilde{\mathcal{I}}$. The resulting tighter relaxation can prevent further branching of the interval containing $\zeta^*_{\text{ub}}(\tilde{\mathcal{I}})$ and leads to better upper bounds since the optimal support from the lower bounding problem is later used to determine the upper bound in the coordinate descent algorithm.

progressively selected and branched upon until either at node j, we have that $g_{\text{lb}}(\mathcal{I}_j)$ is close enough to $g_{\text{ub}}(\mathcal{I}_j)$, or $g_{\text{lb}}(\mathcal{I}_j)$ is larger than $g(\{\hat{\zeta}^*\})$ where $\hat{\zeta}^*$ is the best solution found so far. When no more nodes need to be branched upon, the algorithm can terminate and conclude that $\{\hat{\zeta}^*\}$ is close enough to being globally optimal. Based on $\hat{\zeta}^*$ it is then possible to get a nearly optimal solution $\hat{\boldsymbol{u}}^*$ from problem (7) where ζ is fixed to $\hat{\zeta}^*$. Finally, the exact worst-case CVaR of $\hat{\boldsymbol{u}}^*$ can be obtained by solving problem (7) where $\boldsymbol{u}$ is fixed to $\hat{\boldsymbol{u}}^*$. For clarity, Algorithm 1 presents the pseudocode for the procedure that was described.

Algorithm 1 Spatial branch and bound algorithm for solving problem (7)

procedure SPATIALBRANCH&BOUND(ϵ,n)
  $UB^* \leftarrow \infty$, $\mathcal{N} \leftarrow \{\bar{\mathcal{I}}\}$ ▷ $\mathcal{N}$ denotes the set of nodes
  while $\mathcal{N} \neq \emptyset$ **do**
    Sort $\mathcal{N}$ in non-decreasing order according to $g_{\text{lb}}(\cdot)$
    Take first $\mathcal{I}$ nodes (intervals) out of $\mathcal{N}$
    if $g_{\text{ub}}(\mathcal{I}) < UB^*$ **then**
      $UB^* \leftarrow g_{\text{ub}}(\mathcal{I})$
      $\hat{\zeta}^* \leftarrow \zeta^*_{\text{ub}}(\mathcal{I})$
    end if
    Remove from $\mathcal{N}$, all $\hat{\mathcal{I}}$ such that $g_{\text{lb}}(\hat{\mathcal{I}}) \geq UB^*$
    if $g_{\text{ub}}(\mathcal{I}) > (1+\epsilon) g_{\text{lb}}(\mathcal{I}))$ **then**
      Divide $\mathcal{I}$ into n sub-intervals $\{\mathcal{I}_1, \ldots, \mathcal{I}_n\}$
      $\mathcal{N} \leftarrow \mathcal{N} \cup \{\mathcal{I}_1, \ldots, \mathcal{I}_n\}$
    end if
  end while
  Solve problem (7) with constraint $\zeta = \hat{\zeta}^*$ to get $\hat{\boldsymbol{u}}^*$
  Solve problem (7) with constraint $\boldsymbol{u} = \hat{\boldsymbol{u}}^*$ to get $\hat{t}^*$
  return $\hat{\boldsymbol{u}}^*, \hat{t}^*$
end procedure

We are left with describing how the two operators can be efficiently implemented.

4.1 Using RLT with C&CG for $g_{\text{lb}}(\mathcal{I})$

In this section, we describe an efficient procedure that can be used to establish a lower bound for the optimal value of problem (7). This procedure will need to overcome the two underlying difficulties of problem (7), namely that constraint (4d) is bilinear in $\boldsymbol{u}$ and ζ, and that the size of this problem is exponential with respect to $|E|$ due to the set $\mathcal{L}$. To tackle the first obstacle, we will employ a popular reduced reformulation linearization technique (see Liberti and Pantelides (2006)) that will relax the problem to a linear program. The second obstacle will be dealt with by employing a column generation scheme (Desrosiers and Lübbecke (2005)) that only considers a subset $\hat{\mathcal{L}} \subset \mathcal{L}$ and progressively adds relevant support points to it until optimality conditions are satisfied.

Starting with the idea of relaxing the problem to a linear program, we follow similar steps as used in Liberti and Pantelides (2006). Namely, we start by introducing a set of redundant

constraints in problem (7). This gives rise to the following equivalent optimization model

$$\begin{aligned}
\min_{\substack{\boldsymbol{u},t,\boldsymbol{w}^+,\boldsymbol{w}^-,\beta,\\ \boldsymbol{\eta},\boldsymbol{\theta}^+,\boldsymbol{\theta}^-,\Delta,\zeta}} \quad & t && && \text{(8a)}\\
\text{subject to} \quad & (4c)-(4g),(7b)-(7d), && &&\\
& \sum_{\boldsymbol{\ell}\in\mathcal{L}} \eta_{\boldsymbol{\ell}} = \zeta && && \text{(8b)}\\
& \eta_{\boldsymbol{\ell}} \geq u_{\boldsymbol{\ell}}\zeta_{\text{lb}} && \forall \boldsymbol{\ell}\in\mathcal{L} && \text{(8c)}\\
& \eta_{\boldsymbol{\ell}} \leq u_{\boldsymbol{\ell}}\zeta_{\text{ub}} && \forall \boldsymbol{\ell}\in\mathcal{L} && \text{(8d)}\\
& \eta_{\boldsymbol{\ell}} \geq \zeta + \zeta_{\text{ub}}(u_{\boldsymbol{\ell}}-1) && \forall \boldsymbol{\ell}\in\mathcal{L} && \text{(8e)}\\
& \eta_{\boldsymbol{\ell}} \leq \zeta + \zeta_{\text{lb}}(u_{\boldsymbol{\ell}}-1) && \forall \boldsymbol{\ell}\in\mathcal{L}, && \text{(8f)}\\
& \zeta_{\text{lb}} \leq \zeta \leq \zeta_{\text{ub}}, && && \text{(8g)}
\end{aligned}$$

where ζ_{lb} and ζ_{ub} are the respective boundaries of $\mathcal{I}$. In problem (8), constraint (8b) is redundant since $\sum_{\boldsymbol{\ell}\in\mathcal{L}} u_\ell = 1$ implies that $\sum_{\boldsymbol{\ell}\in\mathcal{L}} u_\ell\zeta = \zeta$, and we have that $u_\ell\zeta = \eta_\ell$. On the other hand, constraints (8c)-(8f) are so-called McCormick inequalities (see McCormick (1976)) which are known to be redundant given that $\zeta \in \mathcal{I}$ and $0 \leq \boldsymbol{u} \leq 1$.

We obtain the linear relaxation of the above problem by removing constraint (4d) from problem (8). For completeness, we present this linear programming relaxation in full details below:

$$\begin{aligned}
\underset{\substack{\boldsymbol{u},\boldsymbol{w}^+,\boldsymbol{w}^-,t,\beta,\\ \boldsymbol{\eta},\boldsymbol{\theta}^+,\boldsymbol{\theta}^-,\Delta,\zeta}}{\text{minimize}} \quad & t && && \text{(9a)}\\
\text{subject to} \quad & \zeta + \Gamma\sqrt{|\mathcal{K}|}\beta + \Gamma\sum_{k'\in\mathcal{K}}(w^+_{k'}+w^-_{k'}) + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}\in\mathcal{L}}\Delta_{\boldsymbol{\ell},k} && &&\\
& +(\hat{\boldsymbol{q}}-\boldsymbol{e}_k)^\top(\boldsymbol{w}^+-\boldsymbol{w}^-+\boldsymbol{\theta}^+-\boldsymbol{\theta}^-) \leq t && \forall k\in\mathcal{K} && \text{(9b)}\\
& \theta^+_k + \theta^-_k - \beta = 0 && \forall k\in\mathcal{K} && \text{(9c)}\\
& \mathbf{1}^\top\boldsymbol{u} = 1 && && \text{(9d)}\\
& \sum_{\boldsymbol{\ell}\in\mathcal{L}}\eta_{\boldsymbol{\ell}} = \zeta && && \text{(9e)}\\
& \eta_{\boldsymbol{\ell}} \geq \zeta + \zeta_{\text{ub}}(u_{\boldsymbol{\ell}}-1) && \forall \boldsymbol{\ell}\in\mathcal{L} && \text{(9f)}\\
& \eta_{\boldsymbol{\ell}} \leq \zeta + \zeta_{\text{lb}}(u_{\boldsymbol{\ell}}-1) && \forall \boldsymbol{\ell}\in\mathcal{L}, && \text{(9g)}\\
& \zeta_{\text{lb}} \leq \zeta \leq \zeta_{\text{ub}} && && \text{(9h)}\\
& \boldsymbol{w}^+ \geq 0, \boldsymbol{w}^- \geq 0, \boldsymbol{\theta}^+ \geq 0,\ \boldsymbol{\theta}^- \geq 0 && && \text{(9i)}\\
& \Delta_{\boldsymbol{\ell},k} \geq u_{\boldsymbol{\ell}} f_{\boldsymbol{\ell},k} - \eta_{\boldsymbol{\ell}} && \forall \boldsymbol{\ell}\in\mathcal{L},\ k\in\mathcal{K} && \text{(9j)}\\
& \Delta_{\boldsymbol{\ell},k} \geq 0 && \forall \boldsymbol{\ell}\in\mathcal{L},\ k\in\mathcal{K} && \text{(9k)}\\
& \boldsymbol{u} \geq 0 && && \text{(9l)}\\
& \eta_{\boldsymbol{\ell}} \geq u_{\boldsymbol{\ell}}\zeta_{\text{lb}} && \forall \boldsymbol{\ell}\in\mathcal{L} && \text{(9m)}\\
& \eta_{\boldsymbol{\ell}} \leq u_{\boldsymbol{\ell}}\zeta_{\text{ub}} && \forall \boldsymbol{\ell}\in\mathcal{L}. && \text{(9n)}
\end{aligned}$$

In what follows, it will be useful to compactly represent (9) as follows:

$$\begin{aligned}
\underset{\boldsymbol{x},\{\boldsymbol{y}_{\boldsymbol{\ell}}\}_{\boldsymbol{\ell}\in\mathcal{L}}}{\text{minimize}} \quad & \boldsymbol{h}^\top\boldsymbol{x} && && \text{(10a)}\\
\text{subject to} \quad & A\boldsymbol{x} + \sum_{\boldsymbol{\ell}\in\mathcal{L}} B_{\boldsymbol{\ell}}\boldsymbol{y}_{\boldsymbol{\ell}} \leq \boldsymbol{s} && && \text{(10b)}\\
& W_{\boldsymbol{\ell}}\boldsymbol{y}_{\boldsymbol{\ell}} \leq \mathbf{0} && \forall \boldsymbol{\ell}\in\mathcal{L}, && \text{(10c)}
\end{aligned}$$

where (10b) summarizes constraints (9b)-(9i) while (10c) summarizes constraints (9j)-(9n). The decision variables are given by

$$\begin{aligned} &\boldsymbol{x} = [(\boldsymbol{w}^+)^\top \ (\boldsymbol{w}^-)^\top \ (\boldsymbol{\theta}^+)^\top \ (\boldsymbol{\theta}^-)^\top \ \zeta \ \beta \ t]^\top, \\ &\boldsymbol{y_\ell} = [\boldsymbol{\Delta}_{\boldsymbol{\ell}}^\top \ u_{\boldsymbol{\ell}} \ \eta_{\boldsymbol{\ell}}]^\top, \qquad\qquad \forall \boldsymbol{\ell} \in \mathcal{L} \end{aligned}$$

where $\boldsymbol{x} \in \mathbb{R}^{4|\mathcal{K}|+3}, \boldsymbol{y_\ell} \in \mathbb{R}^{|\mathcal{K}|+2}$ and $\boldsymbol{\Delta_\ell} \in \mathbb{R}^{|\mathcal{K}|}$ for each $\boldsymbol{\ell} \in \mathcal{L}$, $A \in \mathbb{R}^{(7|\mathcal{K}|+2|\mathcal{L}|+6)\times(4|\mathcal{K}|+3)}$, $B_{\boldsymbol{\ell}} \in \mathbb{R}^{(7|\mathcal{K}|+2|\mathcal{L}|+6)\times(|\mathcal{K}|+2)}$, $W_{\boldsymbol{\ell}} \in \mathbb{R}^{(2|\mathcal{K}|+3)\times(|\mathcal{K}|+2)}$, $\boldsymbol{s} \in \mathbb{R}^{7|\mathcal{K}|+2|\mathcal{L}|+6}$, $\boldsymbol{h} \in \mathbb{R}^{4|\mathcal{K}|+3}$. For the full description of A, B, W, s, h, refer to Appendix B.

The idea behind column generation methods is to consider that at optimality $\boldsymbol{y_\ell} \neq 0$ only for a small set of index $\ell \in \mathcal{L}$. This is a legitimate assumption for our DRMFNI problem where we expect that there should be an optimal strategy that only randomizes among a relatively small (non-exponential) number of interdiction plans. This was observed for instance in the distributionally robust risk neutral facility location problem studied in Delage and Saif (2021).

Given a set $\hat{\mathcal{L}} \subseteq \mathcal{L}$, by linear programming duality, we have that the solution of

$$\begin{aligned} \underset{\boldsymbol{x}, \{\boldsymbol{y_\ell}\}_{\boldsymbol{\ell}\in\mathcal{L}}}{\text{minimize}} \quad & \boldsymbol{h}^\top \boldsymbol{x} && && \text{(11a)} \\ \text{subject to} \quad & A\boldsymbol{x} + \sum_{\boldsymbol{\ell}\in\mathcal{L}} B_{\boldsymbol{\ell}} \boldsymbol{y_\ell} \leq \boldsymbol{s} && && \text{(11b)} \\ & W_{\boldsymbol{\ell}} \boldsymbol{y_\ell} \leq \boldsymbol{0} && \forall \boldsymbol{\ell} \in \hat{\mathcal{L}} && \text{(11c)} \\ & \boldsymbol{y}_\ell = 0 && \forall \ell \in \mathcal{L}/\hat{\mathcal{L}}, && \text{(11d)} \end{aligned}$$

is optimal with respect to problem (10) if and only if a solution of its dual problem

$$\begin{aligned} \underset{\boldsymbol{\psi}, \{\boldsymbol{\sigma_\ell}\}_{\boldsymbol{\ell}\in\hat{\mathcal{L}}}}{\text{maximize}} \quad & -\boldsymbol{\psi}^\top \boldsymbol{s} && && \text{(12a)} \\ \text{subject to} \quad & \boldsymbol{h} + A^\top \boldsymbol{\psi} = \boldsymbol{0} && && \text{(12b)} \\ & B_{\boldsymbol{\ell}}^\top \boldsymbol{\psi} + W_{\boldsymbol{\ell}}^\top \boldsymbol{\sigma_\ell} = \boldsymbol{0} && \forall \boldsymbol{\ell} \in \hat{\mathcal{L}} && \text{(12c)} \\ & \boldsymbol{\psi} \geq 0, \ \boldsymbol{\sigma_\ell} \geq 0 && \forall \boldsymbol{\ell} \in \hat{\mathcal{L}}, && \text{(12d)} \end{aligned}$$

where $\boldsymbol{\psi} \in \mathbb{R}^{7|\mathcal{K}|+2|\mathcal{L}|+6}$ and $\boldsymbol{\sigma_\ell} \in \mathbb{R}^{2|\mathcal{K}|+3}$ are the dual variables associated to constraints (11b) and (11c) respectively, can be completed with some $\boldsymbol{\sigma_\ell} \in \mathbb{R}^{2|\mathcal{K}|+3}$ for all $\boldsymbol{\ell} \in \mathcal{L}/\hat{\mathcal{L}}$ in a way that makes it feasible in the dual of problem (10), i.e. problem (12) where $\hat{\mathcal{L}}$ is replaced with $\mathcal{L}$.

In particular, this can be verified after solving problem (11) for some $\hat{\mathcal{L}} \subseteq \mathcal{L}$ by obtaining a set $(\hat{\boldsymbol{\psi}}, \{\hat{\boldsymbol{\sigma}}_{\boldsymbol{\ell}}\}_{\ell\in\hat{\mathcal{L}}})$ of optimal dual variables for constraints (11b) and (11c) and verifying if they satisfy the following condition:

$$\inf_{\boldsymbol{\ell}\in\mathcal{L}} \sup_{\boldsymbol{\sigma_\ell}\geq 0} \inf_{\boldsymbol{y_\ell}} \hat{\boldsymbol{\psi}}^\top B_{\boldsymbol{\ell}} \boldsymbol{y_\ell} + \boldsymbol{\sigma_\ell}^\top W_{\boldsymbol{\ell}} \boldsymbol{y_\ell} \geq 0. \tag{13}$$

Lemma 4. *Let* $(\hat{\boldsymbol{x}}, \{\hat{\boldsymbol{y}}_{\boldsymbol{\ell}}\}_{\boldsymbol{\ell}\in\mathcal{L}}, \hat{\boldsymbol{\psi}}, \{\hat{\boldsymbol{\sigma}}_{\boldsymbol{\ell}}\}_{\boldsymbol{\ell}\in\hat{\mathcal{L}}})$ *be a primal-dual solution pair for problems* (11) *and* (12) *that satisfies condition* (13)*. Then,* $(\hat{\boldsymbol{x}}, \{\hat{\boldsymbol{y}}_{\boldsymbol{\ell}}\}_{\boldsymbol{\ell}\in\mathcal{L}})$ *is also optimal for problem* (10)*.*

Furthermore, when condition (13) is not satisfied, a violating $\boldsymbol{\ell} \in \mathcal{L}$, which is necessarily not in $\hat{\mathcal{L}}$, can be identified and added to $\hat{\mathcal{L}}$ in order to improve the solution obtained by problem (11).

Two important observations need to be made at this point. First, fortunately enough problem (11) can be shown to reduce to a linear program which size is linear in $|\hat{\mathcal{L}}|$ given that only the decision variables $(\boldsymbol{x}, \{\boldsymbol{y_\ell}\}_{\boldsymbol{\ell}\in\hat{\mathcal{L}}})$ need to be optimized, while the only constraints indexed by some $\boldsymbol{\ell} \in \mathcal{L}/\hat{\mathcal{L}}$ are constraints (11d) and a subset of constraint (11b) which capture constraints (9f)

and (9g). The latter two become redundant for any $\boldsymbol{\ell} \notin \hat{\mathcal{L}}$ since in those cases the constraints reduce to:

$$\zeta_{\text{lb}} \leq \zeta \leq \zeta_{\text{ub}} \qquad \forall \boldsymbol{\ell} \in \mathcal{L}/\hat{\mathcal{L}}.$$

Secondly, one can also show that condition (13) can be verified efficiently by solving a mixed-integer linear program of reasonable size as described in the following proposition.

Proposition 5. *Given some $\mathcal{I} \subseteq \bar{\mathcal{I}}$, verifying condition* (13) *is equivalent to verifying whether the optimal value of the following mixed-integer linear program is non-negative:*

$$\begin{aligned}
\underset{\substack{\boldsymbol{\ell}\in\mathcal{L},\boldsymbol{\Delta},\eta \\ \{\boldsymbol{\lambda}_k,\boldsymbol{\upsilon}_k,\boldsymbol{\Upsilon}_k\}_{k\in\mathcal{K}}}}{\text{minimize}} \quad & \frac{\hat{\boldsymbol{\varphi}}^\top}{1-\alpha}\boldsymbol{\Delta} + \hat{p} + \hat{\pi}\eta && && \text{(14a)}\\
\text{subject to} \quad & \Delta_k \geq \boldsymbol{c}_k^\top(\boldsymbol{\lambda}_k - \boldsymbol{\Upsilon}_k) - \eta && \forall k \in \mathcal{K} && \text{(14b)}\\
& \boldsymbol{\Upsilon}_k \leq \boldsymbol{\ell} && k \in \mathcal{K} && \text{(14c)}\\
& \boldsymbol{\Upsilon}_k \leq \boldsymbol{\lambda}_k && \forall k \in \mathcal{K} && \text{(14d)}\\
& \boldsymbol{\Upsilon}_k \geq \boldsymbol{\lambda}_k + \boldsymbol{\ell} - 1 && \forall k \in \mathcal{K} && \text{(14e)}\\
& \boldsymbol{\Upsilon}_k \geq \mathbf{0} && \forall k \in \mathcal{K} && \text{(14f)}\\
& \boldsymbol{\lambda}_k + N^\top \boldsymbol{\upsilon}_k - \boldsymbol{d} \geq 0 && \forall k \in \mathcal{K} && \text{(14g)}\\
& 0 \leq \boldsymbol{\lambda}_k \leq 1 && \forall k \in \mathcal{K} && \text{(14h)}\\
& \Delta_k \geq 0 && \forall k \in \mathcal{K} && \text{(14i)}\\
& \eta \geq \zeta_{lb} && && \text{(14j)}\\
& \eta \leq \zeta_{ub} && && \text{(14k)}\\
& \mathbf{1}^\top \boldsymbol{\ell} \leq B && && \text{(14l)}\\
& \boldsymbol{\ell} \in \{0,1\}^{|E|}, && && \text{(14m)}\\
& \sum_{i:\bar{\ell}_i=0}(1-\ell_i) + \sum_{i:\bar{\ell}_i=1}\ell_i \leq |E| - 1 && \forall \bar{\ell} \in \{\ell \in \hat{\mathcal{L}} \,|\, \hat{\boldsymbol{y}}_\ell \neq 0\} && \text{(14n)}
\end{aligned}$$

where $\boldsymbol{\lambda}_k \in \mathbb{R}^{|E|}$, $\boldsymbol{\Delta} \in \mathbb{R}^{|\mathcal{K}|}, \eta \in \mathbb{R}$, $\boldsymbol{\upsilon}_k \in \mathbb{R}^{|V|}$ and $\boldsymbol{\Upsilon} \in \mathbb{R}^{|E|\times|\mathcal{K}|}$, $\hat{\boldsymbol{y}}$ solves (11) *for $\hat{\mathcal{L}} \subseteq \mathcal{L}$, while $\hat{\boldsymbol{\varphi}}$, $\hat{p}, \hat{\pi}$ are the elements of $\hat{\boldsymbol{\psi}}$ associated with* (9b), (9d), (9e) *respectively.*

We note that to obtain the reformulation in (14), we exploit the fact that the dual vector (denoted by $\boldsymbol{\lambda}$) associated with constraint (2c) is bounded above by 1. The need for (14n) is also a peculiarity that is due to the fact that constraints (9f) and (9g), indexed by $\boldsymbol{\ell}$, are left in the restricted master problem (11). Furthermore, the size of MILP (14) in Proposition 5 grows linearly with $|\hat{\mathcal{L}}|$ where $\hat{\mathcal{L}}$ is a smaller set compared to $\mathcal{L}$ for our DRMFNI problem if the optimal randomized strategy can be supported on a small number of interdiction plans. For completeness, we present the pseudocode of the column generation algorithm described above in Algorithm 2. Note that the algorithm is guaranteed to converge in a finite number of iterations given that at each iteration, either the algorithm terminates or a new element $\boldsymbol{\ell} \in \mathcal{L}$ is added to $\hat{\mathcal{L}}$, yet $|\mathcal{L}|$ is finite.

4.2 Using coordinate descent for $g_{\mathbf{ub}}(\mathcal{I})$

Given an interval $\mathcal{I} \subseteq \bar{\mathcal{I}}$ we are looking for an upper bound on $g(\mathcal{I})$ and a value of $\zeta \in \mathcal{I}$ such that $g(\mathcal{I})$ matches this upper bound. In order to accomplish this task, we will first look back at the solution of problem (9) to identify the optimal support $\mathcal{L}^*_{\text{lb}}$ and distribution $\boldsymbol{u}^*_{\text{lb}}$ of the lower bounding problem. We then perform coordinate descent on problem (7) where $\mathcal{L}$ is replaced with $\mathcal{L}^*_{\text{lb}}$ iterating between a step where $\boldsymbol{u}$ stays fixed at the best solution found so far, initially

Algorithm 2 Column Generation Algorithm to solve problem (9)

procedure ColumnGeneration
  $\hat{\mathcal{L}} \leftarrow \{\mathbf{0}\}$
  while $\hat{\mathcal{L}} \neq \mathcal{L}$ **do**
    Solve problem (9) with $\eta_{\boldsymbol{\ell}} = 0$,$u_{\boldsymbol{\ell}} = 0$, $\boldsymbol{\Delta}_{\boldsymbol{\ell}} = \mathbf{0}$ for all $\boldsymbol{\ell} \in \mathcal{L}/\hat{\mathcal{L}}$
    Identify a set of optimal dual variables $\hat{\boldsymbol{\varphi}}$, $\hat{p}$, and $\hat{\pi}$ as defined in Proposition 5.
    Solve problem (14) to obtain optimal value v^* and optimal $\boldsymbol{\ell}^*$
    if $v^* \geq 0$ **then**
      return Optimal solution obtained in step 4
    else
      $\hat{\mathcal{L}} \leftarrow \hat{\mathcal{L}} \cup \{\boldsymbol{\ell}^*\}$
    end if
  end while
  return Solve (9) and return optimal solution
end procedure

at $\boldsymbol{u}^*_{\text{lb}}$, and a step where it is rather ζ that stays fixed. In both cases, the problem reduces to a linear program whose size is linear in the size of $\mathcal{L}^*_{\text{lb}}$. This procedure is considered to have converged when the relative improvement on optimal value is considered small enough. For completeness, we provide the pseudocode for the coordinate descent algorithm in Algorithm 3.

Algorithm 3 Coordinate Descent Algorithm to obtain upper bound on problem (7)

procedure CoordinateDescent(ϵ, $\mathcal{L}^*_{\text{lb}}$, $\boldsymbol{u}^*_{\text{lb}}$)
  $\bar{\boldsymbol{u}}^* \leftarrow \boldsymbol{u}^*_{\text{lb}}$, $\mathcal{L} \leftarrow \mathcal{L}^*_{\text{lb}}$
  do
    Solve problem (7) with $\boldsymbol{u} = \bar{\boldsymbol{u}}^*$ to get optimal value t^*_1 and optimal $\zeta^*_{\text{ub}}(\mathcal{I})$
    Solve problem (7) with $\zeta = \zeta^*_{\text{ub}}(\mathcal{I})$ to get optimal value t^*_2 and optimal $\bar{\boldsymbol{u}}^*$
  while $t^*_2 < (1-\epsilon)t^*_1$
  $g_{\text{ub}}(\mathcal{I}) \leftarrow t^*_1$
  return $g_{\text{ub}}(\mathcal{I})$, $\zeta^*_{\text{ub}}(\mathcal{I})$
end procedure

5 Numerical experiments

We performed a series of numerical experiments to show the convergence and numerical efficiency of the spatial branch and bound algorithm (in Section 5.1), to quantify the value of risk aversion modeling in Section 5.2, and to compare the performance of our optimal randomized strategies, Loizou's randomized strategies (Loizou (2015)), and deterministic interdiction plans in Section 5.3. All algorithms were implemented in Matlab 2020a using the YALMIP toolbox and CPLEX 12.9.0 was used to solve all continuous and mixed-integer linear programs, except in Section 5.1.3 where we compare our algorithm with Gurobi's bilinear solver and therefore, we solve all continuous, mixed integer linear programs, and bilinear programs using Gurobi 9.1.2. The Matlab codes used to generate the results can be found on GitHub[3].

[3]GitHub repository: https://github.com/Utsav19/Value-of-Randomization

With the exception of Section 5.1.3, all DRMFNI problem instances were generated using the grid network structure given in Figure 1. Within the same column, the arcs can point upward or downward with equal probability whereas between different columns, the arcs always point in the direction of the sink. The number of rows is denoted by m and the number of columns is denoted by n. This class of network instances is the same as the one used in Cormican et al. (1998), Royset and Wood (2007), Janjarassuk and Linderoth (2008) and Atamtürk et al. (2020) where the arcs in the first and last column and those leaving the source node $\mathfrak{s}$ or entering the sink $\mathfrak{t}$ are not interdictable and have infinite capacity. For each DRMFNI problem instance, capacity vector scenarios are drawn from a factor model, $\boldsymbol{c} := F\boldsymbol{\xi}$, with each ξ_i independently distributed according to an exponential distribution with mean μ_i, for some fixed $F \in \mathbb{R}_+^{|E|\times 2}$ and $\boldsymbol{\mu} \in \mathbb{R}_+^2$ that were randomly generated for the given instance. The empirical distribution over the $|\mathcal{K}|$ i.i.d. observations is used as a reference distribution for our ambiguity set $\mathcal{Q}$, i.e. $\hat{\boldsymbol{q}} = (1/|\mathcal{K}|)\mathbf{1}$, as is commonly done in DRO (see Ben-Tal et al. (2013), Bayraksan and Love (2015), Mohajerin Esfahani and Kuhn (2018), Lam (2019), Ji and Lejeune (2021), Duchi et al. (2021)).

We note that in these experiments, we will pay special attention to the Value of the Randomized Strategy (VRS), which can be defined as the relative gap between the worst-case CVaR obtained by the deterministic and randomized strategies:

$$\text{VRS} = \frac{\max_{\boldsymbol{q}\in\hat{\mathcal{Q}}_{|\mathcal{K}|}} \text{CVaR}^{\alpha}_{k\sim\boldsymbol{q}}[f_{\hat{\boldsymbol{\ell}}^*_{|\mathcal{K}|},k}] - \max_{\boldsymbol{q}\in\hat{\mathcal{Q}}_{|\mathcal{K}|}} \text{CVaR}^{\alpha}_{\boldsymbol{\ell}\sim\hat{\boldsymbol{u}}^*_{|\mathcal{K}|},k\sim\boldsymbol{q}}[f_{\boldsymbol{\ell},k}]}{\max_{\boldsymbol{q}\in\hat{\mathcal{Q}}_{|\mathcal{K}|}} \text{CVaR}^{\alpha}_{\boldsymbol{\ell}\sim\hat{\boldsymbol{u}}^*_{|\mathcal{K}|},k\sim\boldsymbol{q}}[f_{\boldsymbol{\ell},k}]} \times 100\%.$$

where $\hat{\boldsymbol{\ell}}^*_{|\mathcal{K}|}$ and $\hat{\boldsymbol{u}}^*_{|\mathcal{K}|}$ are respectively the optimal deterministic and randomized solutions of the DRMFNI problem, and where $\hat{\mathcal{Q}}_{|\mathcal{K}|}$ is centered at the empirical distribution over the sample set.

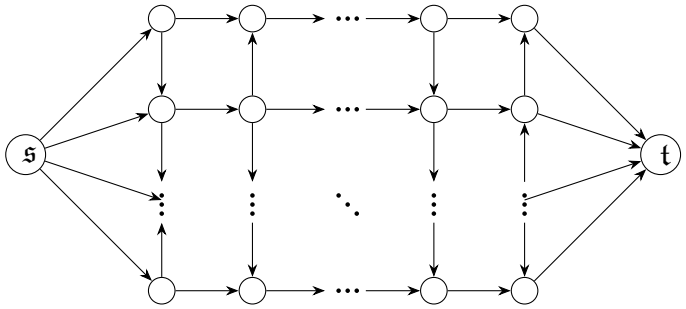

Figure 1: Grid network for numerical experiments

5.1 Computational efficiency

Next, we report the computation times of our spatial branch and bound algorithm in Section 5.1.1, and compare it with a heuristic approach based on solving DRMFNI problem for different fixed values of ζ (in Section 5.1.2). A comparison of the performance of our spatial branch and bound algorithm and Gurobi's bilinear solver is finally given in Section 5.1.3.

5.1.1 Convergence of spatial branch and bound algorithm.

In this subsection, we show that the spatial branch and bound algorithm described in Section 4 converges quickly to approximately optimal solutions. The size of the instances is given by $m \times n \times |\mathcal{K}|$ where m and n denote the number of rows and columns, respectively, in the network and $|\mathcal{K}|$ denotes the number of scenarios. We randomly generate 10 instances of sizes

$10 \times 10 \times 10$, $15 \times 15 \times 15$, and $20 \times 20 \times 20$ of the DRMFNI problem as described in the introduction of this section. The scenarios are randomly generated from the constructed factor model. For $\Gamma = 0$, our problem reduces to a CVaR minimization with probability equal to the reference probability (e.g. , uniform) in which case deterministic plans are optimal since CVaR is a mixture quasiconcave risk measure (a.k.a. randomization proof), see Delage et al. (2019).

We then solve the instances for a convergence tolerance, ϵ, either equal to 0.01% or 0.0001%, terminating the algorithm if the higher precision is not achieved after 2 hours. The results of the numerical experiments for $B = \lfloor 0.3m \rfloor$, $\alpha = 0.05$ are given in Table 2 where $\lfloor i \rfloor$ denotes the largest integer less than or equal to i. We find that for networks of all sizes, the spatial branch and bound algorithm converges within 2 hours to an accuracy of $\epsilon = 0.0001\%$ for all 10 problem instances.

Table 2: Computation times of spatial branch and bound algorithm for different levels of uncertainty (Γ) for 10 randomly generated instances.

Size ($m \times n \times \lvert\mathcal{K}\rvert$)	Type of performance	Level of Uncertainty (Γ) (in % of $\lvert\mathcal{K}\rvert$)		
		5%	10%	15%
$10 \times 10 \times 10$	avg. cpu time (s) ($\epsilon = 0.01\%$)	1.34	14.83	10.69
	avg. cpu time (s) ($\epsilon = 0.0001\%$)	1.40	17.29	15.19
$15 \times 15 \times 15$	avg. cpu time (s) ($\epsilon = 0.01\%$)	16.53	24.59	21.20
	avg. cpu time (s) ($\epsilon = 0.0001$)%	25.81	45.30	42.31
$20 \times 20 \times 20$	avg. cpu time (s) ($\epsilon = 0.01\%$)	10.75	12.25	11.37
	avg. cpu time (s) ($\epsilon = 0.0001\%$)	22.76	24.60	22.30

In producing Table 2, we observed that for most of the instances used to produce Table 2, the VRS is found to be equal to 0. Therefore, we identify a subclass of instances with VRS $\neq 0$ to show that our algorithm also converges for those instances. For each network of size $10 \times 10 \times 10$, $15 \times 15 \times 15$, and $20 \times 20 \times 20$, we sampled scenarios from the factor model until we have obtained 10 instances with VRS $> 0.1\%$ for each value of Γ in $\{0.05|\mathcal{K}|, 0.1|\mathcal{K}|, 0.15|\mathcal{K}|\}$. The results are given in Table 3 for $B = \lfloor 0.3m \rfloor$ and $\alpha = 0.05$. We find that the spatial branch and bound algorithm converges within 2 hours with a tolerance of 0.0001% for all instances of sizes $10 \times 10 \times 10$, $15 \times 15 \times 15$ and $20 \times 20 \times 20$. Finally, we refer interested reader to Section F in the Appendix, which provides empirical evidence that the choice of reference distribution does not affect computation time significantly.

Table 3: Computation times of spatial branch and bound algorithm for different levels of uncertainty (Γ) for 10 randomly generated instances for which $VRS > 0.1\%$.

Size ($m \times n \times \lvert\mathcal{K}\rvert$)	Type of performance	Level of Uncertainty (Γ) (in % of $\lvert\mathcal{K}\rvert$)		
		5%	10%	15%
$10 \times 10 \times 10$	avg. cpu time (s) ($\epsilon = 0.01\%$)	66.21	44.28	56.35
	avg. cpu time (s) ($\epsilon = 0.0001\%$)	212.21	245.14	193.39
$15 \times 15 \times 15$	avg. cpu time (s) ($\epsilon = 0.01\%$)	34.05	55.19	36.84
	avg. cpu time (s) ($\epsilon = 0.0001$)%	151.97	114.88	220.46
$20 \times 20 \times 20$	avg. cpu time (s) ($\epsilon = 0.01\%$)	141.64	96.81	152.71
	avg. cpu time (s) ($\epsilon = 0.0001$)%	591.81	195.11	393.83

5.1.2 Comparison of spatial branch and bound algorithm with heuristic approach.

In this section, we compare the efficiency of our spatial branch and bound algorithm with a heuristic approach that is based on the observation that for a fixed ζ, (7) reduces to a linear program.

Table 4: Computation times of spatial branch and bound algorithm for different levels of uncertainty (Γ) for 10 randomly generated instances of problem size ($10 \times 10 \times 10$) with precision $\epsilon = 0.01\%$ with the heuristic

Partitions N	Type of performance	Instances in Table 2 Level of Uncertainty (Γ) (in % of $\lvert\mathcal{K}\rvert$)			Instances in Table 3 Level of Uncertainty (Γ) (in % of $\lvert\mathcal{K}\rvert$)		
		5%	10%	15%	5%	10%	15%
4	avg. cpu time (s)	3.03	2.4	2.11	7.77	9.61	5.6
	optimality gap	1.24%	1.64%	1.64%	1.01%	1.12%	1.21%
16	avg. cpu time (s)	11.97	9.41	7.91	29.29	30.03	19.49
	optimality gap	0.27%	0.34%	0.34%	0.21%	0.32%	0.21%
32	avg. cpu time (s)	21.83	20.76	19.13	41.16	63.77	35.12
	optimality gap	0.16%	0.15%	0.15%	0.1%	0.13%	0.14%
64	avg. cpu time (s)	48.20	38.1	29.66	129.17	118.69	78.64
	optimality gap	0.06%	0.09%	0.09%	0.06%	0.06%	0.05%
128	avg. cpu time (s)	98.80	66.72	61.63	257.43	213.96	155
	optimality gap	0.03%	0.04%	0.04%	0.04%	0.02%	0.03%
spatial B&B	avg. cpu time (s)	1.34	14.83	10.69	66.21	44.28	56.35
	optimality gap	0%	$< 10^{-5}\%$	$< 10^{-5}\%$	$< 10^{-3}\%$	$< 10^{-4}\%$	$< 10^{-3}\%$

The heuristic can be summarized as follows. First, we choose a uniform grid on the interval $[\zeta_{\text{lb}}, \zeta_{\text{ub}}]$ to create N partitions. Next, for each ζ value on the grid, we solve problem (7) to obtain an upper bound on the interdictor's DRMFNI problem. The best upper bound (feasible solution) is the one with the minimum value among the $N+1$ solutions corresponding to each ζ on the grid. We use a column generation procedure to solve problem (7) for a fixed ζ. In Table 4, we report the optimality gap for the best feasible solutions identified by using this heuristic approach and the spatial branch and bound algorithm with precision $\epsilon = 0.01\%$ relative to the lower bound identified using the spatial branch and bound algorithm with $\epsilon = 0.0001\%$. The results are given for instances considered in Tables 2 and 3 with a network of size $10 \times 10 \times 10$, $B = 3$, $\hat{\boldsymbol{q}} = (1/10)\mathbf{1}$, $\alpha = 0.05$ and $N \in \{4, 16, 32, 64, 128\}$. The spatial branch and bound algorithm achieves less than $10^{-3}\%$ optimality gap on average for all values of Γ while the heuristic approach takes significantly higher time than our algorithm to reach an optimality gap of around 0.05%. This implies that our column generation scheme identifies near optimal support for the original DRMFNI of the interdictor even when the convex relaxation is not tight. As a result, the subsequent iterations of the spatial branch and bound algorithm after optimality gap of $\epsilon = 0.01\%$ has been achieved improve the lower bound to close the gap between the lower and upper bounds.

5.1.3 Comparison of Algorithm 1 with Gurobi.

Next, we compare the performance of Algorithm 1 with Gurobi's bilinear solver on smaller instances. To solve problem (7), we consider two variants of Algorithm 1, one in which we use column generation procedure described in Algorithm 2 to solve the lower bounding problem (9), and another in which we directly feed problem (9) to an LP solver. In addition to a $5 \times 5 \times |\mathcal{K}|$ grid network, the computational experiments are conducted on two real-world networks, namely, Sioux-Falls (LeBlanc et al. 1975) and NOBEL-US (Orlowski et al. 2010), that have been used

in the literature (c.f. Lei et al. (2018), Song and Shen (2016)). The source and sink nodes are taken as 1 and 24, respectively for the Sioux-Falls network (see Figure 4 in Section E of the appendix), and for the NOBEL-US network (see Figure 5 in Section E of the appendix), Seattle and Princeton are taken as the source and sink nodes, respectively .

Table 5 reports the average computation times to solve 10 instances of problem (7) for each of the three network instances with $\alpha = 0.05$, $\Gamma = 0.05|\mathcal{K}|$, $|\mathcal{K}| \in \{100, 150, 200\}$, using Gurobi's bilinear solver and the two variants of Algorithm 1 which are described above.

We can see that, unlike for Algorithm 1, Gurobi's bilinear solver's solution time does not scale well with the total number of possible interdiction plans. This is even the case when comparing to the version of spatial B&B that does not employ column generation. For instance, Gurobi's bilinear solver does not find a feasible solution for the NOBEL-US network (861 possible plans) with 200 scenarios within 10 minutes while Algorithm 1 converges with an optimality gap of 0.01% even without using the column generation procedure. For Sioux-Falls network with a small number of interdiction plans, we find that it is better to use Algorithm 1 without the CG procedure because the small size of the problem does not justify the computation time to solve the MILP within the CG procedure.

Table 5: Comparison of average computation times for the Gurobi's bilinear solver and our spatial branch and bound algorithm described in Algorithm 1 with- and without column generation for 10 instances and a convergence tolerance of 0.01%.

	$\lvert\mathcal{L}\rvert$	$\lvert\mathcal{K}\rvert$	Gurobi (s)	spatial B&B (s)	spatial B&B + CG (s)
Sioux-Falls $\lvert E\rvert = 76$, $\lvert V\rvert = 24$, $B = 1$	76	100	0.71	0.95	2.42
	76	150	1.25	1.38	3.66
	76	200	2.23	1.72	6.66
Grid Network $\lvert E\rvert = 50$, $\lvert V\rvert = 27$, $B = 2$	496	100	31.54	35.87	7.16
	496	150	67.64*	31.16	10.47
	496	200	73.70	43.32	13.25
Nobel-US $\lvert E\rvert = 42$, $\lvert V\rvert = 14$, $B = 2$	861	100	49.56	24.85	7.94
	861	150	200.10	20.12	5.52
	861	200	N.A.**	30.30	8.34

* average computed on nine instances as one failed to converge in 10 min
**all 10 instances failed to converge in 10 min

5.2 Value of modeling risk-aversion in DRMFNI model

In this section, we quantify the importance of accounting for risk aversion in the DRFMNI problem. In particular, we evaluate the worst-case CVaR achieved by an optimal interdiction strategy from the risk-neutral DRFMNI model (i.e. when $\alpha = 0\%$) and compare it to the minimal worst-case CVaR that can be reached. We consider the 10 random instances of size $10 \times 10 \times 10$ studied in Table 3 for $\Gamma = 1$ and $B = 3$. First, we solve the risk-neutral model, and compute the optimal randomized strategy ($\hat{u}^0$) for the interdictor. For the randomized strategy ($\hat{u}^0$), we determine the worst-case CVaR for the risk-averse DRMFNIP for $\alpha \in \{0.5, 0.1, 0.2, 0.3, 0.4\}$. The relative gap between the worst-case CVaR for $\hat{u}^0$ and optimal worst-case CVaR for risk-averse DRMFNI problem is defined as the value of risk-averse model (VRAM) and is given by

$$\text{VRAM} = \frac{\max_{\boldsymbol{q} \in \hat{\mathcal{Q}}_{|\mathcal{K}|}} \text{CVaR}^{\alpha}_{\boldsymbol{\ell} \sim \hat{\boldsymbol{u}}^0_{|\mathcal{K}|}, k \sim \boldsymbol{q}}[f_{\boldsymbol{\ell},k}] - \max_{\boldsymbol{q} \in \hat{\mathcal{Q}}_{|\mathcal{K}|}} \text{CVaR}^{\alpha}_{\boldsymbol{\ell} \sim \hat{\boldsymbol{u}}^*_{|\mathcal{K}|}, k \sim \boldsymbol{q}}[f_{\boldsymbol{\ell},k}]}{\max_{\boldsymbol{q} \in \hat{\mathcal{Q}}_{|\mathcal{K}|}} \text{CVaR}^{\alpha}_{\boldsymbol{\ell} \sim \hat{\boldsymbol{u}}^*_{|\mathcal{K}|}, k \sim \boldsymbol{q}}[f_{\boldsymbol{\ell},k}]} \times 100\%,$$

where $\hat{\boldsymbol{u}}^*_{|\mathcal{K}|}$ is an optimal solution to the DRFMNI with risk aversion α. It can be seen in Figure 2 that the value of risk-averse model increases with risk-aversion parameter α. For e.g. we observed at an $\alpha = 40\%$ level, that the sub-optimality of the solution of the risk neutral DRMFNI could reach up to nearly 11% when compared to the minimum worst-case conditional value-at-risk. This evidence therefore supports the claim that risk aversion is an important element to model even in an environment with distribution ambiguity.

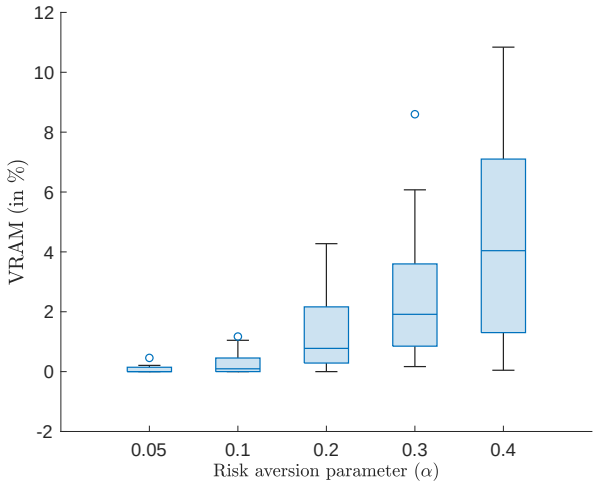

Figure 2: Box plots summarizing the statistics (quartiles and outliers) of the value of risk-averse model (VRAM) as a function of the risk aversion level.

5.3 Value of Randomization

Next, in Section 5.3.1, we compare the in-sample value of randomized strategies for our model and Loizou's model with respect to the optimal deterministic strategies for the instances generated for the grid network in Figure 1. In Section 5.3.2, we provide evidence that our randomized strategies can outperform deterministic strategies in out-of-sample experiments.

5.3.1 Performance of our randomized, Loizou's randomized, and deterministic strategies

To illustrate the benefit of randomization, we generate instances of the network given in Figure 1 for 10 rows and 10 columns. We are interested in comparing the performance of strategies that are obtained using only $|\mathcal{K}| = 10$ scenarios from the underlying distribution. To do so, we generated 100 set of samples for the capacities of the arcs for each level of $\Gamma \in \{0.5,\ 1,\ 1.5\}$. For each set of samples, we compare the performance of the randomized interdiction strategy $\hat{\boldsymbol{u}}^*_{10}$ and the deterministic strategy $\hat{\ell}^*_{10}$ that optimizes the DRMFNI problem constructed from this "observed" sample set, with $B \in \{3,\ 6,\ 9\}$, $\alpha = \{0.05,\ 0.1,\ 0.2,\ 0.4\}$, and $\hat{\boldsymbol{q}} = (1/10)\mathbf{1}$. The optimal randomized strategy $\hat{\boldsymbol{u}}^*_{10}$ is obtained by using the spatial branch and bound algorithm outlined in Section 4 with $\epsilon = 0.01\%$, while the optimal deterministic strategy $\hat{\boldsymbol{\ell}}^*_{10}$ is computed by solving the MILP given in Section C in the appendix.

To compute the value of randomized strategies from Loizou (2015), we solve the following distributionally robust network interdiction problem where the interdictor minimizes the worst-

case CVaR of the expected maximum flow over her own randomized strategy:

$$\text{(L-DRMFNI)} \quad \underset{\boldsymbol{u}\in\Delta\mathcal{L}}{\text{minimize}} \max_{\boldsymbol{q}\in\mathcal{Q}} \inf_{\zeta} \; \zeta + \frac{1}{1-\alpha}\sum_{k} q_k \left[\left(\sum_{\boldsymbol{\ell}} u_{\boldsymbol{\ell}} f_{\boldsymbol{\ell},k}\right) - \zeta\right]^{+}, \tag{15}$$

The robust counterpart of the L-DRMFNI problem is a linear program which is solved by using our column generation algorithm.

Table 6 reports the number of instances, for each level of Γ, for which VRS was between zero and 0.1%, between 0.1% and 1%, and above 1% for L-DRMFNI and DRMFNI models. Also, it can be shown that L-DRMFNI always underestimates the risk associated with the DRMFNI model (Delage et al. 2019, Proposition 3). Therefore, the VRS for L-DRMFNI problem is always greater than or equal to the VRS for the DRMFNI problem which is confirmed by our numerical experiments. In particular, L-DRMFNI mislabeled 11.22% (i.e. 404 of the 3600 instances) of our randomly generated instances as being instances with a strictly positive VRS.

We observe that for higher budget of interdiction, the randomized strategies perform significantly better than deterministic ones with respect to DRMFNI problem instances, reaching for $B = 9$ an average VRS of 11.66% when $\alpha = 0.05$. We argue that this evidence confirms that there is a real observable benefit for the network interdictor for employing randomized interdiction plans in a risk averse network interdiction problem, both in a distributionally robust setting (c.f. the performance comparison) and in a setting where the network capacity distribution information comes from a limited number of observed realizations. Also, as the distributional ambiguity increases from $\Gamma = 1$ to $\Gamma = 1.5$, there is no further improvement in performance for a variation in either the budget of interdiction B or the risk aversion parameter α.

In Table 6, we can see that there is no variation in VRS with α for $\Gamma = 1$ and $\Gamma = 1.5$ for L-DRMFNI problem because we have considered 10 scenarios and the model reduces to minimizing the worst-case expected payoff similar to the model in Bertsimas et al. (2016).

5.3.2 Out-of-sample performance of randomized and deterministic strategies

In this subsection, we compare the out-of-sample performance of optimal randomized and deterministic strategies of the interdictor with respect to the DRMFNI problem.

For a budget of interdiction, $B = 9$, and risk aversion parameter, $\alpha = 0.1$, we found in the previous section that there are 15 network instances for which VRS $\geq 1\%$ for all values of $\Gamma \in \{0.5, 1, 1.5\}$. For these 17 network instances, we conduct the training and test experiments by randomly sampling probability distributions over the capacities of the arcs from a Dirichlet distribution, with parameters $\beta_j = \beta$ for all $j = 1, \ldots, N$ where $\beta \in \{0.1, 0.3, 0.5, 0.8\}$. For each instance, we generate 100 distributions for the in-sample study, and 10000 distributions for the out-of-sample experiments. We chose the Γ values in such a way that the Euclidean ball of radius Γ centred at the uniform distribution contains 95% of the probability distributions in the training sample. We solve the DRMFNI problem to obtain optimal deterministic and randomized strategies and then, use them to compute the corresponding out-of-sample CVaR.

Table 7 reports the in-sample average value of randomized strategies for DRMFNI model for $\alpha = 0.1$. We can see that as the concentration parameter increases, the VRS decreases because larger values of β result in less uncertainty about the distribution.

In Figure 3, the box plots for the $95th$ percentile of the out-of-sample CVaR corresponding to the optimal deterministic and randomized strategies are given for different values of the concentration parameter. We can see from the box plots that randomized strategies outperform the deterministic plans for all values of the concentration parameter (β), and for lower values of β, the difference in performance is significant.

Table 6: Value of randomized solution performances from our DRMFNI and L-DRMFNI (Loizou (2015)) models for problem of size $10 \times 10 \times 10$.

Level of Uncertainty (Γ)	B	α	VRS based on DRMFNI				VRS based on L-DRMFNI			
			[0% 0.1%)	[0.1%, 1%)	$\geq 1\%$		[0% 0.1%]	[0.1%, 1%)	$\geq 1\%$	
			# inst.	# inst.	# inst.	avg. VRS	# inst.	# inst.	# inst.	avg. VRS
5%	3	0.05	94	3	3	2.39%	90	5	5	2.40%
		0.1	95	3	2	2.09%	89	7	4	2.80%
		0.2	96	4	0	0%	87	5	8	2.17%
		0.4	98	2	0	0%	85	4	11	3.37%
	6	0.05	85	6	9	3.56%	78	9	13	3.78%
		0.1	88	5	7	3.25%	76	11	13	3.94%
		0.2	93	4	3	3.43%	76	5	19	3.43%
		0.4	90	8	2	1.59%	69	3	28	4.44%
	9	0.05	71	9	20	6.58%	67	4	29	6.98%
		0.1	80	5	15	6.22%	66	5	29	7.59%
		0.2	87	2	11	4.66%	65	2	33	8.13%
		0.4	93	1	6	4.58%	62	4	34	13.01%
10%	3	0.05	87	5	8	3.30%	85	4	11	3.37%
		0.1	88	5	7	2.81%	85	4	11	3.37%
		0.2	91	5	4	1.85%	85	4	11	3.37%
		0.4	98	2	0	0%	85	4	11	3.37%
	6	0.05	69	8	23	3.95%	69	3	28	4.44%
		0.1	71	9	20	3.42%	69	3	28	4.44%
		0.2	80	7	13	2.86%	69	3	28	4.44%
		0.4	91	7	2	1.57%	69	3	28	4.44%
	9	0.05	67	3	30	11.66%	62	4	34	13.01%
		0.1	71	2	27	9.92%	62	4	34	13.01%
		0.2	79	4	17	9.22%	62	4	34	13.01%
		0.4	94	0	6	4.38%	62	4	34	13.01%
15%	3	0.05	87	5	8	3.30%	85	4	11	3.37%
		0.1	88	5	7	2.81%	85	4	11	3.37%
		0.2	91	5	4	1.86%	85	4	11	3.37%
		0.4	98	2	0	0%	85	4	11	3.37%
	6	0.05	69	8	23	3.95%	69	3	28	4.44%
		0.1	71	9	20	3.42%	69	3	28	4.44%
		0.2	80	7	13	2.86%	69	3	28	4.44%
		0.4	91	7	2	1.57%	69	3	28	4.44%
	9	0.05	67	3	30	11.66%	62	4	34	13.01%
		0.1	71	2	27	9.92%	62	4	34	13.01%
		0.2	79	4	17	9.22%	62	4	34	13.01%
		0.4	94	0	6	4.38%	62	4	34	13.01%

Table 7: In-sample average value of randomized strategy from DRMFNI model for problem of size $10 \times 10 \times 10$ for different values of the concentration parameter β of the Dirichlet distribution

β	avg. VRS
0.1	15.40%
0.3	10.71%
0.5	7.22%
0.8	5.57%

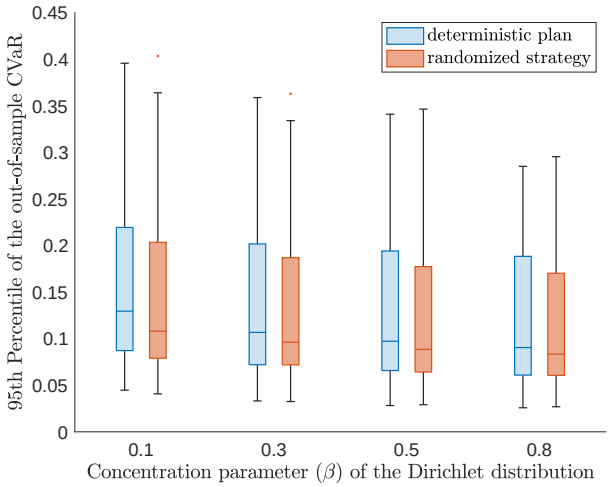

Figure 3: Statistics (over 15 problem instances) of the 95th percentile of the out-of-sample CVaR for optimal deterministic plans and randomized strategies for different values of the concentration parameter of the Dirichlet distribution

6 Conclusions

In this paper, we introduced a distributionally robust risk averse maximum flow network interdiction problem to model the strategic interactions between a risk-averse interdictor and the flow player. We solved the interdictor's bilinear DRO problem by first reformulating it as a bilinear optimization problem using LP duality and then devising a spatial branch and bound algorithm. After observing that the optimal randomized strategy can be supported on a small number of interdiction plans in DRO problems, we developed a column generation algorithm that can be used to efficiently determine the convex relaxation of the interdictor's problem. Our numerical experiments showed that 1) our proposed spatial branch and bound algorithm can efficiently solve distributionally robust interdiction problems of reasonable sizes; 2) randomization can be quite effective in reducing the risk exposure obtained from the optimal deterministic interdiction strategy when comparing the worst-case CVaR performances. We note that our spatial branch and bound algorithm does not exploit the specific structure of the max-flow problem and thus should be adaptable to other Stackelberg games where the followers' problem takes the form of a linear program. There is also a possibility that future research might identify special acceleration schemes for this max-flow problems.

Given that Stackelberg games with single leader and multiple followers have been extensively applied in the literature, it would be interesting as future work to extend our algorithm in a way that can address Stackelberg games with a leader that is both risk and ambiguity averse while followers implement a Nash equilibrium that accounts for their respective risk aversion. Another direction for future research could be to determine the value of randomization in distributionally robust shortest path network interdiction problems with a risk averse interdictor. Another problem that can be of interest is to combine our column generation algorithm that scales with the number of interdiction plans with a constraint generation procedure (see Section D in the appendix) in order to potentially handle a much larger number of scenarios for the arcs' capacities.

Acknowledgement

The first author's research is supported by GERAD Postdoctoral Fellowship and FRQNT Postdoctoral research scholarship [Grant 275296 and 301065]. The second author acknowledges the support from the Canadian Natural Sciences and Engineering Research Council and the Canada Research Chair program [Grant RGPIN-2016-05208 and 492997-2016].

A Proofs

A.1 Proof of Proposition 1

We substitute the expression of CVaR in (3) to obtain

$$\underset{\boldsymbol{u}\in\Delta\mathcal{L}}{\text{minimize}} \max_{\boldsymbol{q}\in\mathcal{Q}} \inf_{\zeta} \quad \zeta + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k u_{\boldsymbol{\ell}}[f_{\boldsymbol{\ell},k} - \zeta]^+. \tag{16}$$

We prove by contradiction that an optimal ζ, denoted by ζ^*, lies between 0 and $\bar{\zeta}$. First, we assume that $\zeta^* = \zeta_L < 0$ is the largest optimal value for ζ. The maximum flow $f_{\boldsymbol{\ell},k}$ for any $\boldsymbol{\ell}$ and k is bounded below by 0, hence for any $\boldsymbol{q} \in \mathcal{Q}$ the CVaR equals

$$\zeta_L + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k u_{\boldsymbol{\ell}}(f_{\boldsymbol{\ell},k} - \zeta_L) = \frac{(1-\alpha)\zeta_L - \zeta_L}{1-\alpha} + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k u_{\boldsymbol{\ell}} f_{\boldsymbol{\ell},k}.$$

However, we arrive at a contradiction since $\zeta = 0$ is at least as good as ζ_L:

$$\frac{-\alpha\zeta_L}{1-\alpha} + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k u_{\boldsymbol{\ell}} f_{\boldsymbol{\ell},k} \geq \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k u_{\boldsymbol{\ell}} f_{\boldsymbol{\ell},k} = 0 + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k u_{\boldsymbol{\ell}}[f_{\boldsymbol{\ell},k} - 0]^+,$$

for any $\alpha \in [0, 1)$. So, we can conclude that $\zeta^* := 0$ is also optimal.

Alternatively, we can assume that $\zeta^* = \zeta_H > \bar{\zeta}$ is the smallest optimal solution for ζ. In this case, for any $\boldsymbol{q} \in \mathcal{Q}$ we have that CVaR equals:

$$\zeta_H + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k u_{\boldsymbol{\ell}}[f_{\boldsymbol{\ell},k} - \zeta_H]^+ = \zeta_H,$$

since $f_{\boldsymbol{\ell},k} \leq f_{\mathbf{0},k} \leq \bar{\zeta} < \zeta_H$. Yet, for $\zeta = \bar{\zeta}$, the worst-case CVaR, given by $\bar{\zeta}$, is strictly less than ζ_H which contradicts our assumption that ζ_H is the smallest optimal solution for ζ.

We now turn ourselves to establishing the reformulation presented as problem (4). Let $v(\zeta, \boldsymbol{q}, \boldsymbol{u}) = \zeta + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k u_{\boldsymbol{\ell}}[f_{\boldsymbol{\ell},k} - \zeta]^+$. Since $v(\zeta, \boldsymbol{q}, \boldsymbol{u})$ is convex in ζ for all $\boldsymbol{q} \in \mathcal{Q}$ while being linear in $\boldsymbol{q}$ for all $\zeta \in [0, \bar{\zeta}]$, and since $[0, \bar{\zeta}]$ is bounded and $\mathcal{Q}$ is convex, it follows from Sion's minimax theorem (see Sion (1958)) that (16) is equivalent to

$$\underset{\boldsymbol{u}\in\Delta\mathcal{L}}{\text{minimize}} \min_{0\leq\zeta\leq\bar{\zeta}} \max_{\boldsymbol{q}\in\mathcal{Q}} \quad \zeta + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k u_{\boldsymbol{\ell}}[f_{\boldsymbol{\ell},k} - \zeta]^+. \tag{17}$$

Since we are minimizing over $\boldsymbol{u}$ and ζ, we have an equivalent reformulation of (17) given by

$$\begin{aligned} \underset{\substack{\boldsymbol{u}\in\Delta\mathcal{L},\boldsymbol{\Delta}\geq 0,\ \boldsymbol{q}\in\mathcal{Q} \\ 0\leq\zeta\leq\bar{\zeta}}}{\text{minimize}} \max \quad & \zeta + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}}\sum_{k} q_k \Delta_{\boldsymbol{\ell},k} \\ \text{subject to} \quad & \Delta_{\boldsymbol{\ell},k} \geq u_{\boldsymbol{\ell}} f_{\boldsymbol{\ell},k} - u_{\boldsymbol{\ell}}\zeta \qquad\qquad \forall \boldsymbol{\ell} \in \mathcal{L},\ k \in \mathcal{K}. \end{aligned}$$

Employing an epigraph representation of the above objective function and introducing the decision variable $\eta_{\boldsymbol{\ell}} = u_{\boldsymbol{\ell}}\zeta$, we obtain problem (4). □

A.2 Proof of Proposition 2

We start with the following lemma which applied the theory proposed in Ben-Tal et al. (2015) to derive reformulations of non-linear robust constraints.

Lemma 6. *Suppose Assumption 1 holds. Then* $(\zeta, \boldsymbol{\Delta}, t)$ *satisfy the robust constraint* (4b) *if and only if there exists a vector* $\boldsymbol{\gamma} \in \mathbb{R}^{|\mathcal{K}|}$ *that satisfies the following constraint*

$$\hat{\boldsymbol{q}}^\top \boldsymbol{\gamma} + \delta^*(\boldsymbol{\gamma}|\mathcal{Z}) - g_*(\zeta, \Delta, t, \boldsymbol{\gamma}) \leq 0 \tag{18}$$

where $\delta^*(\boldsymbol{\gamma}|\mathcal{Z})$ *denotes the support function of the set of perturbations* $\mathcal{Z}$ *and* $g_*(\zeta, \Delta, t, \boldsymbol{\gamma})$ *is the partial concave conjugate of* $g(\zeta, \Delta, t, \boldsymbol{q})$ *where*

$$g(\zeta, \Delta, t, \boldsymbol{q}) = \zeta + \frac{1}{1-\alpha} \sum_{\boldsymbol{\ell} \in \mathcal{L}} \sum_{k \in \mathcal{K}} q_k \Delta_{\boldsymbol{\ell}, k} - t$$

with domain $\{\boldsymbol{q} | \boldsymbol{q} \geq 0,\ \sum_k q_k\ = 1\}$.

Proof. Proof of Lemma 6 We first rewrite constraint (4b) as:

$$h(\zeta, \Delta, t, \boldsymbol{z}) \leq 0 \quad \forall \boldsymbol{z} \in \mathcal{Z}$$

where $h(\zeta, \Delta, t, \boldsymbol{z}) := g(\zeta, \Delta, t, \hat{\boldsymbol{q}} + \boldsymbol{z})$. Clearly, $h(\zeta, \Delta, t, \boldsymbol{z})$ is affine in $\boldsymbol{z}$ for all (ζ, Δ, t), and the intersection of $\mathcal{Z}$ and the domain of $h(\zeta, \Delta, t, \boldsymbol{z})$, namely $\mathcal{Q}$, has a non-empty relative interior as long as $\Gamma > 0$ (e.g. with $\hat{\boldsymbol{z}} = 0$ when $\hat{\boldsymbol{q}} = 1/|\mathcal{K}|$. Thus, Fenchel duality theorem can be used to show that constraint (4b) is equivalent to (see Ben-Tal et al. (2015)):

$$\delta^*(\boldsymbol{\gamma}|\mathcal{Z}) - h_*(\zeta, \Delta, t, \boldsymbol{\gamma}) \leq 0\,.$$

Furthermore, one can show that:

$$h_*(\zeta, \Delta, t, \boldsymbol{\gamma}) = \inf_{\boldsymbol{z}} \boldsymbol{\gamma}^\top \boldsymbol{z} - g(\zeta, \Delta, t, \hat{\boldsymbol{q}} + \boldsymbol{z}) = \inf_{\boldsymbol{q}} \boldsymbol{\gamma}^\top (\boldsymbol{q} - \hat{\boldsymbol{q}}) - g(\zeta, \Delta, t, \boldsymbol{q}) = -\boldsymbol{\gamma}^\top \hat{\boldsymbol{q}} + g_*(\zeta, \Delta, t, \boldsymbol{\gamma})\,. \square$$

$\square$

Lemma 7. *The partial concave conjugate function of* $g(\zeta, \Delta, t, \boldsymbol{q})$ *is given by*

$$g_*(\zeta, \Delta, t, \boldsymbol{\gamma}) = \min_{k=1,2,\ldots,|\mathcal{K}|} \gamma_k - \zeta - \frac{1}{1-\alpha} \sum_{\boldsymbol{\ell} \in \mathcal{L}} \Delta_{\boldsymbol{\ell}, k} + t \tag{19}$$

Proof. Proof of Lemma 7

The conjugate function of $g(\zeta, \Delta, t, \boldsymbol{q})$ is given by

$$\begin{aligned} g_*(\zeta, \Delta, t, \boldsymbol{\gamma}) &= \inf_{\boldsymbol{q}} \boldsymbol{\gamma}^\top \boldsymbol{q} - g(\zeta, \Delta, t, \boldsymbol{\gamma}) \\ &= \min_{\boldsymbol{q}: \boldsymbol{q} \geq 0,\ \sum_k q_k\ = 1} \sum_{k \in \mathcal{K}} \gamma_k q_k - \zeta + t - \frac{1}{1-\alpha} \sum_{\boldsymbol{\ell} \in \mathcal{L}} \sum_{k \in \mathcal{K}} q_k \Delta_{\boldsymbol{\ell}, k} \\ &= \min_{k=1,2,\ldots,K} \gamma_k - \zeta + t - \frac{1}{1-\alpha} \sum_{\boldsymbol{\ell} \in \mathcal{L}} \Delta_{\boldsymbol{\ell}, k} \end{aligned}$$

where the last equality follows from the fact that the minimization is of an affine function over a bounded polyhedron, thus the minimum achieved at an extreme point.$\square$ $\square$

With the two lemmas in hand, it is straightforward to obtain constraint (5b) as an equivalent reformulation of (4b). $\square$

A.3 Proof of Proposition 3

The support function of $\mathcal{Z}$ is given by

$$\delta^*(\boldsymbol{\gamma}|\mathcal{Z}) = \sup_{\boldsymbol{z}\in\mathcal{Z}} \boldsymbol{\gamma}^\top \boldsymbol{z} \tag{20}$$

An equivalent representation of set $\mathcal{Z}(\Gamma)$ is given by

$$\mathcal{Z}(\Gamma) = \left\{ \boldsymbol{z} \in \mathbb{R}^{|\mathcal{K}|} \;\middle|\; \exists \boldsymbol{\delta} \in \mathbb{R}^{|\mathcal{K}|}, \begin{array}{ll} -\Gamma \le z_k \le \Gamma & \forall k \in \mathcal{K} \\ -\delta_k \le z_k \le \delta_k & \forall k \in \mathcal{K} \\ \sum_{k\in\mathcal{K}} \delta_k = \Gamma\sqrt{|\mathcal{K}|} & \end{array} \right\}.$$

Therefore, we can rewrite the maximization problem in (20) as follows:

$$\begin{array}{lll} \underset{\boldsymbol{\delta},\boldsymbol{z}}{\text{maximize}} & \displaystyle\sum_{k\in\mathcal{K}} \gamma_k z_k & & \text{(21a)} \\ \text{subject to} & z_k \le \Gamma & \forall k \in \mathcal{K} & \text{(21b)} \\ & z_k \ge -\Gamma & \forall k \in \mathcal{K} & \text{(21c)} \\ & z_k \le \delta_k & \forall k \in \mathcal{K} & \text{(21d)} \\ & z_k \ge -\delta_k & \forall k \in \mathcal{K} & \text{(21e)} \\ & \displaystyle\sum_{k\in\mathcal{K}} \delta_k = \Gamma\sqrt{|\mathcal{K}|} & & \text{(21f)} \end{array}$$

Since $\boldsymbol{z} = 0$ is feasible, we can use strong LP duality to obtain the support function of $\mathcal{Z}$:

$$\begin{array}{lll} \delta^*(\boldsymbol{\gamma}|\mathcal{Z}) = \underset{\substack{\boldsymbol{w}^+,\boldsymbol{w}^-,\boldsymbol{\theta}^+,\\ \boldsymbol{\theta}^-,\beta,}}{\text{minimize}} & \Gamma\sqrt{|\mathcal{K}|}\beta + \Gamma\displaystyle\sum_{k\in\mathcal{K}}(w_k^+ + w_k^-) & \\ \quad \text{subject to} & \gamma_k - w_k^+ + w_k^- - \theta_k^+ + \theta_k^- = 0 & \forall k \in \mathcal{K} \\ & \theta_k^+ + \theta_k^- - \beta = 0 & \forall k \in \mathcal{K} \\ & \boldsymbol{w}^+ \ge 0,\, \boldsymbol{w}^- \ge 0,\, \boldsymbol{\theta}^+ \ge 0,\, \boldsymbol{\theta}^- \ge 0, & \end{array}$$

where $\boldsymbol{w}^+ \in \mathbb{R}^{|\mathcal{K}|}$, $\boldsymbol{w}^- \in \mathbb{R}^{|\mathcal{K}|}$, $\boldsymbol{\theta}^+ \in \mathbb{R}^{|\mathcal{K}|}$, $\boldsymbol{\theta}^- \in \mathbb{R}^{|\mathcal{K}|}$, and $\beta \in \mathbb{R}$ are the dual variables associated to constraints (21b) to (21f), respectively.

Combining the above problem with (5), we obtain

$$\begin{array}{lll} \underset{\substack{\boldsymbol{u},t,\boldsymbol{\gamma},\boldsymbol{w}^+,\boldsymbol{w}^-,\beta\\ \boldsymbol{\eta},\boldsymbol{\theta}^+,\boldsymbol{\theta}^-,\Delta,\zeta}}{\text{minimize}} & t & & \text{(22a)} \\ \quad \text{subject to} & \zeta + \Gamma\sqrt{|\mathcal{K}|}\beta + \Gamma\displaystyle\sum_{k'\in\mathcal{K}}(w_{k'}^+ + w_{k'}^-) + \frac{1}{1-\alpha}\sum_{\boldsymbol{\ell}\in\mathcal{L}}\Delta_{\boldsymbol{\ell},k} & & \\ & \quad + \hat{\boldsymbol{q}}^\top\boldsymbol{\gamma} - \gamma_k \le t & \forall k \in \mathcal{K} & \text{(22b)} \\ & \gamma_k - w_k^+ + w_k^- - \theta_k^+ + \theta_k^- = 0 & \forall k \in \mathcal{K} & \text{(22c)} \\ & \theta_k^+ + \theta_k^- - \beta = 0 & \forall k \in \mathcal{K} & \text{(22d)} \\ & \boldsymbol{w}^+ \ge 0,\, \boldsymbol{w}^- \ge 0,\, \boldsymbol{\theta}^+ \ge 0,\, \boldsymbol{\theta}^- \ge 0 & & \text{(22e)} \\ & \text{(4c)} - \text{(4h)}. & & \text{(22f)} \end{array}$$

Since $\gamma_k = w_k^+ - w_k^- + \theta_k^+ - \theta_k^-$ for all $k \in \mathcal{K}$, an equivalent representation of the above problem

is given by

$$\begin{aligned}
\underset{\substack{\boldsymbol{u},t,\boldsymbol{w}^+,\boldsymbol{w}^-,\beta\\ \boldsymbol{\eta},\boldsymbol{\theta}^+,\boldsymbol{\theta}^-,\Delta,\zeta}}{\text{minimize}} \quad & t \\
\text{subject to} \quad & \text{(4c)} - \text{(4h)},\ \text{(22d)} - \text{(22e)} \\
& \zeta + \Gamma\sqrt{|\mathcal{K}|}\beta + \Gamma \sum_{k' \in \mathcal{K}} (w_{k'}^+ + w_{k'}^-) + \frac{1}{1-\alpha} \sum_{\boldsymbol{\ell} \in \mathcal{L}} \Delta_{\boldsymbol{\ell},k} \\
& + \sum_{k' \in \mathcal{K}} (\hat{\boldsymbol{q}} - \boldsymbol{e}_k)^\top (\boldsymbol{w}^+ - \boldsymbol{w}^- + \boldsymbol{\theta}^+ - \boldsymbol{\theta}^-) \leq t \qquad \forall k \in \mathcal{K}. \square
\end{aligned}$$

A.4 Proof of Lemma 4

This proof relies on confirming that when condition (13) is satisfied, we can construct an assignment $\{\bar{\boldsymbol{\sigma}}_{\boldsymbol{\ell}}\}_{\boldsymbol{\ell} \in \mathcal{L}}$ such that $(\hat{\boldsymbol{x}}, \{\hat{\boldsymbol{y}}_{\boldsymbol{\ell}}\}_{\boldsymbol{\ell} \in \mathcal{L}}, \hat{\boldsymbol{\psi}}, \{\bar{\boldsymbol{\sigma}}_{\boldsymbol{\ell}}\}_{\boldsymbol{\ell} \in \mathcal{L}})$ is a valid primal-dual solution pair for problem (10). Specifically, if condition (13) is satisfied, then for all $\boldsymbol{\ell} \in \mathcal{L}$, there must exist a $\bar{\boldsymbol{\sigma}}_{\boldsymbol{\ell}} \geq 0$ such that $\inf_{\boldsymbol{y}_{\boldsymbol{\ell}}} \hat{\boldsymbol{\psi}}^\top B_{\boldsymbol{\ell}} \boldsymbol{y}_{\boldsymbol{\ell}} + \bar{\boldsymbol{\sigma}}_{\boldsymbol{\ell}}^\top W_{\boldsymbol{\ell}} \boldsymbol{y}_{\boldsymbol{\ell}} \geq 0$, which implies that $B_{\boldsymbol{\ell}}^\top \hat{\boldsymbol{\psi}} + W_{\boldsymbol{\ell}}^\top \bar{\boldsymbol{\sigma}}_{\boldsymbol{\ell}} = \boldsymbol{0}$. Hence, this provides a recipe for assembling some $(\hat{\boldsymbol{\psi}}, \{\bar{\boldsymbol{\sigma}}_{\boldsymbol{\ell}}\}_{\boldsymbol{\ell} \in \mathcal{L}})$ that satisfies all constraints of the dual of problem (10). Since $(\hat{\boldsymbol{x}}, \{\hat{\boldsymbol{y}}_{\boldsymbol{\ell}}\}_{\boldsymbol{\ell} \in \mathcal{L}}, \hat{\boldsymbol{\psi}}, \{\hat{\boldsymbol{\sigma}}_{\boldsymbol{\ell}}\}_{\boldsymbol{\ell} \in \hat{\mathcal{L}}})$ is a primal-dual solution of problem (11), we must have that $\boldsymbol{h}^\top \hat{\boldsymbol{x}} = -\hat{\boldsymbol{\psi}}^\top \boldsymbol{s}$. Hence, $(\hat{\boldsymbol{x}}, \{\hat{\boldsymbol{y}}_{\boldsymbol{\ell}}\}_{\boldsymbol{\ell} \in \mathcal{L}}, \hat{\boldsymbol{\psi}}, \{\bar{\boldsymbol{\sigma}}_{\boldsymbol{\ell}}\}_{\boldsymbol{\ell} \in \mathcal{L}})$ is a valid primal-dual solution pair for problem (10). $\square$

A.5 Proof of Proposition 5

We start by repeating the definition of condition (13):

$$\inf_{\boldsymbol{\ell} \in \mathcal{L}} \sup_{\boldsymbol{\sigma}_{\boldsymbol{\ell}} \geq 0} \inf_{\boldsymbol{y}_{\boldsymbol{\ell}}} \hat{\boldsymbol{\psi}}^\top B_{\boldsymbol{\ell}} \boldsymbol{y}_{\boldsymbol{\ell}} + \boldsymbol{\sigma}_{\boldsymbol{\ell}}^\top W_{\boldsymbol{\ell}} \boldsymbol{y}_{\boldsymbol{\ell}} \geq 0.$$

We will first argue that the order of $\sup_{\boldsymbol{\sigma}_{\boldsymbol{\ell}} \geq 0}$ and $\inf_{\boldsymbol{y}_{\boldsymbol{\ell}}}$ can be changed without affecting the value that is obtained. In particular,

$$\sup_{\boldsymbol{\sigma}_{\boldsymbol{\ell}} \geq 0} \inf_{\boldsymbol{y}_{\boldsymbol{\ell}}} \hat{\boldsymbol{\psi}}^\top B_{\boldsymbol{\ell}} \boldsymbol{y}_{\boldsymbol{\ell}} + \boldsymbol{\sigma}_{\boldsymbol{\ell}}^\top W_{\boldsymbol{\ell}} \boldsymbol{y}_{\boldsymbol{\ell}} = \sup_{\boldsymbol{\sigma}_{\boldsymbol{\ell}} \geq 0 : B_{\boldsymbol{\ell}}^\top \hat{\boldsymbol{\psi}} + W_{\boldsymbol{\ell}}^\top \boldsymbol{\sigma}_{\boldsymbol{\ell}} = 0} 0,$$

while

$$\inf_{\boldsymbol{y}_{\boldsymbol{\ell}}} \sup_{\boldsymbol{\sigma}_{\boldsymbol{\ell}} \geq 0} \hat{\boldsymbol{\psi}}^\top B_{\boldsymbol{\ell}} \boldsymbol{y}_{\boldsymbol{\ell}} + \boldsymbol{\sigma}_{\boldsymbol{\ell}}^\top W_{\boldsymbol{\ell}} \boldsymbol{y}_{\boldsymbol{\ell}} = \inf_{\boldsymbol{y}_{\boldsymbol{\ell}} : W_{\boldsymbol{\ell}} \boldsymbol{y}_{\boldsymbol{\ell}} \leq 0} \hat{\boldsymbol{\psi}}^\top B_{\boldsymbol{\ell}} \boldsymbol{y}_{\boldsymbol{\ell}}.$$

Since the two problems are dual of each other and $\inf_{\boldsymbol{y}_{\boldsymbol{\ell}} : W_{\boldsymbol{\ell}} \boldsymbol{y}_{\boldsymbol{\ell}} \leq 0} \hat{\boldsymbol{\psi}}^\top B_{\boldsymbol{\ell}} \boldsymbol{y}_{\boldsymbol{\ell}}$ is feasible with $\boldsymbol{y}_{\boldsymbol{\ell}} = 0$, we can conclude by strong LP duality that the two values are the same.

We thus obtain that the left-hand side of condition (13) can be obtained by solving:

$$\underset{\boldsymbol{\ell} \in \mathcal{L}, \boldsymbol{y}_{\boldsymbol{\ell}}}{\text{minimize}} \sup_{\boldsymbol{\sigma}_{\boldsymbol{\ell}} \geq 0} \hat{\boldsymbol{\psi}}^\top B_{\boldsymbol{\ell}} \boldsymbol{y}_{\boldsymbol{\ell}} + \boldsymbol{\sigma}_{\boldsymbol{\ell}}^\top W_{\boldsymbol{\ell}} \boldsymbol{y}_{\boldsymbol{\ell}}.$$

On taking the dual of the inner maximization problem, we have

$$\begin{aligned}
\underset{\boldsymbol{\ell} \in \mathcal{L}, \boldsymbol{y}_{\boldsymbol{\ell}}}{\text{minimize}} \quad & \hat{\boldsymbol{\psi}}^\top B_{\boldsymbol{\ell}} \boldsymbol{y}_{\boldsymbol{\ell}} \\
\text{subject to} \quad & W_{\boldsymbol{\ell}} \boldsymbol{y}_{\boldsymbol{\ell}} \leq 0,
\end{aligned}$$

which can be written more carefully as

$$\begin{aligned}
\underset{\boldsymbol{\ell},\bar{\boldsymbol{\Delta}},\bar{u},\bar{\eta}}{\text{minimize}} \quad & \frac{\hat{\boldsymbol{\varphi}}^\top}{1-\alpha}\bar{\boldsymbol{\Delta}} + \hat{p}\bar{u} + \hat{\pi}\bar{\eta} + \hat{\psi}_{1,\ell}(\zeta_{\text{ub}}\bar{u} - \bar{\eta}) + \hat{\psi}_{2,\ell}(\bar{\eta} - \zeta_{\text{lb}}\bar{u}) && && \text{(23a)}\\
\text{subject to} \quad & \bar{\Delta}_k \geq \bar{u} f_{\boldsymbol{\ell},k} - \bar{\eta} && \forall k \in \mathcal{K} && \text{(23b)}\\
& \bar{\Delta}_k \geq 0 && \forall k \in \mathcal{K} && \text{(23c)}\\
& \bar{u} \geq 0 && && \text{(23d)}\\
& \bar{\eta} \geq \bar{u}\zeta_{\text{lb}} && && \text{(23e)}\\
& \bar{\eta} \leq \bar{u}\zeta_{\text{ub}} && && \text{(23f)}\\
& \boldsymbol{\ell} \in \{0,1\}^{|E|} && && \text{(23g)}\\
& \mathbf{1}^\top \boldsymbol{\ell} \leq B, && && \text{(23h)}
\end{aligned}$$

where $\bar{\boldsymbol{\Delta}} \in \mathbb{R}^{|\mathcal{K}|}, \bar{u} \in \mathbb{R}, \bar{\eta} \in \mathbb{R}$, and $\hat{\boldsymbol{\varphi}}$, $\hat{p}$, $\hat{\pi}$, $\hat{\psi}_{1,\ell}$, and $\hat{\psi}_{2,\ell}$ are the terms of the dual vector $\hat{\boldsymbol{\psi}}$ associated with constraints (9b), (9d), (9e), (9f), and (9g) respectively. Note that we can without loss of generality assume that $\hat{\psi}_{1,\ell} = \hat{\psi}_{2,\ell} = 0$ for all $\ell \notin \hat{\mathcal{L}}$ since constraints (9f) and (9g) are considered redundant in the master problem. Moreover, this can also be assumed the case for all $\ell \in \hat{\mathcal{L}}$ such that $\hat{y}_\ell = 0$ in the master problem. Finally, for $\ell \in \{\ell \in \hat{\mathcal{L}} \,|\, \hat{y}_\ell \neq 0\}$, condition (13) is necessarily satisfied since such $\ell \in \hat{\mathcal{L}}$. We are therefore left with condition (13) reducing to the non-negativity of the minimum of the following optimization problem:

$$\begin{aligned}
\underset{\boldsymbol{\ell},\bar{\boldsymbol{\Delta}},\bar{u},\bar{\eta}}{\text{minimize}} \quad & \frac{\hat{\boldsymbol{\varphi}}^\top}{1-\alpha}\bar{\boldsymbol{\Delta}} + \hat{p}\bar{u} + \hat{\pi}\bar{\eta} && && \text{(24a)}\\
\text{subject to} \quad & \bar{\Delta}_k \geq \bar{u} f_{\boldsymbol{\ell},k} - \bar{\eta} && \forall k \in \mathcal{K} && \text{(24b)}\\
& \bar{\Delta}_k \geq 0 && \forall k \in \mathcal{K} && \text{(24c)}\\
& \bar{u} \geq 0 && && \text{(24d)}\\
& \bar{\eta} \geq \bar{u}\zeta_{\text{lb}} && && \text{(24e)}\\
& \bar{\eta} \leq \bar{u}\zeta_{\text{ub}} && && \text{(24f)}\\
& \boldsymbol{\ell} \in \{0,1\}^{|E|} && && \text{(24g)}\\
& \mathbf{1}^\top \boldsymbol{\ell} \leq B, && && \text{(24h)}\\
& \ell \notin \{\ell \in \hat{\mathcal{L}} \,|\, \hat{y}_\ell \neq 0\} && && \text{(24i)}
\end{aligned}$$

The last constraint (24i) can be represented using a list of linear inequalities:

$$\sum_{i:\bar{\ell}_i=0} (1 - \ell_i) + \sum_{i:\bar{\ell}_i=1} \ell_i \leq |E| - 1 \;,\; \forall \bar{\ell} \in \{\ell \in \hat{\mathcal{L}} \,|\, \hat{y}_\ell \neq 0\}.$$

Next, we can observe that when $\bar{u} = 0$, problem (24) necessarily evaluates to zero. From this we conclude that $\bar{u} > 0$ can be added to problem (24) without affecting the conclusion when used to check condition (13). Moreover, $\bar{u}$ can be pulled out of (24) after replacing $\boldsymbol{\Delta} := (1/\bar{u})\bar{\boldsymbol{\Delta}}$

and $\eta := \bar{\eta}/\bar{u}$ to obtain:

$$
\begin{aligned}
\underset{\boldsymbol{\ell},\boldsymbol{\Delta},\eta}{\text{minimize}} \quad & \frac{\hat{\boldsymbol{\varphi}}^\top}{1-\alpha}\boldsymbol{\Delta} + \hat{p} + \hat{\pi}\eta && && \text{(25a)}\\
\text{subject to} \quad & \Delta_k \geq f_{\boldsymbol{\ell},k} - \eta && \forall k \in \mathcal{K} && \text{(25b)}\\
& \Delta_k \geq 0 && \forall k \in \mathcal{K} && \text{(25c)}\\
& \eta \geq \zeta_{\text{lb}} && && \text{(25d)}\\
& \eta \leq \zeta_{\text{ub}} && && \text{(25e)}\\
& \boldsymbol{\ell} \in \{0,1\}^{|E|} && && \text{(25f)}\\
& \mathbf{1}^\top \boldsymbol{\ell} \leq B && && \text{(25g)}\\
& \sum_{i:\bar{\ell}_i=0} (1-\ell_i) + \sum_{i:\bar{\ell}_i=1} \ell_i \leq |E| - 1 && \forall \bar{\ell} \in \{\ell \in \hat{\mathcal{L}} \,|\, y^*_\ell \neq 0\} && \text{(25h)}
\end{aligned}
$$

Next, in order to solve the problem (25), we need to make explicit the relation between $\boldsymbol{\ell}$ and $f_{\boldsymbol{\ell},k}$ for each scenario k. One way is to exploit the dual problem associated with problem (2) which is given by

$$
\begin{aligned}
f_{\boldsymbol{\ell},k} = \quad \min_{\boldsymbol{\upsilon},\boldsymbol{\lambda}} \quad & (\mathbf{1}-\boldsymbol{\ell})^\top C^k \boldsymbol{\lambda}\\
\text{subject to} \quad & \boldsymbol{\lambda} + N^\top \boldsymbol{\upsilon} - \boldsymbol{d} \geq 0\\
& 0 \leq \boldsymbol{\lambda} \leq 1,
\end{aligned}
$$

where $\boldsymbol{\upsilon} \in \mathbb{R}^{|V|}$ and $\boldsymbol{\lambda} \in \mathbb{R}^{|E|}$ are duals associated with the constraints (2b) and (2c) respectively. The dual vector $\boldsymbol{\lambda}$ is bounded above by 1 since a unit increase in capacity of an arc can increase the flow by at most one unit, see, (Cormican et al. 1998, Lemma 1).

We therefore have reduced the evaluation of the left-hand side of condition (13) to solving the following non-linear mixed integer programming (NL-MIP) problem

$$
\begin{aligned}
\underset{\substack{\boldsymbol{\ell},\boldsymbol{\Delta},\eta\\ \{\boldsymbol{\lambda}_k,\boldsymbol{\upsilon}_k\}_{k\in\mathcal{K}}}}{\text{minimize}} \quad & \frac{\hat{\boldsymbol{\varphi}}^\top}{1-\alpha}\boldsymbol{\Delta} + \hat{p} + \hat{\pi}\eta && && \text{(27a)}\\
\text{subject to} \quad & \Delta_k \geq \mathbf{1}^\top C^k \boldsymbol{\lambda}_k - \boldsymbol{\ell}^\top C^k \boldsymbol{\lambda}_k - \eta && \forall k \in \mathcal{K} && \text{(27b)}\\
& \boldsymbol{\lambda}_k + N^\top \boldsymbol{\upsilon}_k - \boldsymbol{d} \geq 0 && \forall k \in \mathcal{K} && \text{(27c)}\\
& 0 \leq \boldsymbol{\lambda}_k \leq 1 && \forall k \in \mathcal{K} && \text{(27d)}\\
& \Delta_k \geq 0 && \forall k \in \mathcal{K} && \text{(27e)}\\
& \eta \geq \zeta_{\text{lb}} && && \text{(27f)}\\
& \eta \leq \zeta_{\text{ub}} && && \text{(27g)}\\
& \boldsymbol{\ell} \in \{0,1\}^{|E|} && && \text{(27h)}\\
& \mathbf{1}^\top \boldsymbol{\ell} \leq B. && && \text{(27i)}\\
& \sum_{i:\bar{\ell}_i=0} (1-\ell_i) + \sum_{i:\bar{\ell}_i=1} \ell_i \leq |E| - 1 && \forall \bar{\ell} \in \{\ell \in \hat{\mathcal{L}} \,|\, \hat{y}_\ell \neq 0\} && \text{(27j)}
\end{aligned}
$$

The non-linearity in the above problem is due to the bilinear terms $\boldsymbol{\ell}^\top C_k \boldsymbol{\lambda}_k = \boldsymbol{c}_k^\top \operatorname{diag}(\boldsymbol{\ell})\boldsymbol{\lambda}_k$.

We can linearize them to obtain an equivalent MILP since $\boldsymbol{\ell}$ is binary:

$$
\begin{aligned}
\underset{\substack{\boldsymbol{\ell},\boldsymbol{\Delta},\eta \\ \{\boldsymbol{\lambda}_k,\boldsymbol{\upsilon}_k,\boldsymbol{\Upsilon}_k\}_{k\in\mathcal{K}}}}{\text{minimize}} \quad & \frac{\hat{\boldsymbol{\varphi}}^\top}{1-\alpha}\boldsymbol{\Delta} + \hat{p} + \hat{\pi}\eta & \\
\text{subject to} \quad & \Delta_k \geq \boldsymbol{c}_k^\top \boldsymbol{\lambda}_k - \boldsymbol{c}_k^\top \boldsymbol{\Upsilon}_k - \eta & \forall k \in \mathcal{K} \\
& \boldsymbol{\Upsilon}_k \leq \boldsymbol{\ell} & k \in \mathcal{K} \\
& \boldsymbol{\Upsilon}_k \leq \boldsymbol{\lambda}_k & \forall k \in \mathcal{K} \\
& \boldsymbol{\Upsilon}_k \geq \boldsymbol{\lambda}_k + \boldsymbol{\ell} - 1 & \forall k \in \mathcal{K} \\
& \boldsymbol{\Upsilon}_k \geq \mathbf{0} & \forall k \in \mathcal{K} \\
& (27c) - (27j), &
\end{aligned}
$$

where each $\boldsymbol{\Upsilon}_k \in \mathbb{R}^{|E|}$ is a linearization of $\operatorname{diag}(\boldsymbol{\ell})\boldsymbol{\lambda}_k$. $\square$

B Matrices

The coefficient matrices in (10) are given by:

$$
\boldsymbol{h} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ 1 \end{pmatrix}, \quad W_{\boldsymbol{\ell}} = \begin{pmatrix} -\mathbb{I} & \boldsymbol{f}_{\boldsymbol{\ell}} & -\mathbf{1} & \text{(9j)} \\ -\mathbb{I} & \mathbf{0} & \mathbf{0} & \text{(9k)} \\ 0 & -1 & 0 & \text{(9l)} \\ 0 & \zeta_{\text{lb}} & -1 & \text{(9m)} \\ 0 & -\zeta_{\text{ub}} & 1 & \text{(9n)} \end{pmatrix},
$$

$$
A = \begin{pmatrix}
\Gamma\mathbf{1}\mathbf{1}^\top - \mathbb{I} + \mathbf{1}\hat{\boldsymbol{q}}^\top & \Gamma\mathbf{1}\mathbf{1}^\top - \mathbf{1}\hat{\boldsymbol{q}}^\top + \mathbb{I} & \mathbf{1}\hat{\boldsymbol{q}}^\top - \mathbb{I} & \mathbb{I} - \mathbf{1}\hat{\boldsymbol{q}}^\top & \mathbf{1} & \Gamma\sqrt{|\mathcal{K}|}\mathbf{1} & -\mathbf{1} & \text{(9b)} \\
0 & 0 & \mathbb{I} & \mathbb{I} & 0 & -\mathbf{1} & 0 & \text{(9c)} \\
0 & 0 & -\mathbb{I} & -\mathbb{I} & 0 & \mathbf{1} & 0 & \text{(9c)} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \text{(9d)} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \text{(9d)} \\
0 & 0 & 0 & 0 & -1 & 0 & 0 & \text{(9e)} \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & \text{(9e)} \\
0 & 0 & 0 & 0 & \mathbf{1} & 0 & 0 & \text{(9f)} \\
0 & 0 & 0 & 0 & -\mathbf{1} & 0 & 0 & \text{(9g)} \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & \text{(9h)} \\
0 & 0 & 0 & 0 & -1 & 0 & 0 & \text{(9h)} \\
-\mathbb{I} & 0 & 0 & 0 & 0 & 0 & 0 & \text{(9i)} \\
0 & -\mathbb{I} & 0 & 0 & 0 & 0 & 0 & \text{(9i)} \\
0 & 0 & -\mathbb{I} & 0 & 0 & 0 & 0 & \text{(9i)} \\
0 & 0 & 0 & -\mathbb{I} & 0 & 0 & 0 & \text{(9i)}
\end{pmatrix},
$$

$$B_{\boldsymbol{\ell}} = \begin{pmatrix} \frac{\mathbb{I}}{1-\alpha} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \\ 0 & \zeta_{\text{ub}}\boldsymbol{e}_{\boldsymbol{\ell}} & -\boldsymbol{e}_{\boldsymbol{\ell}} \\ 0 & -\zeta_{\text{lb}}\boldsymbol{e}_{\boldsymbol{\ell}} & \boldsymbol{e}_{\boldsymbol{\ell}} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; \boldsymbol{s} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ \zeta_{\text{ub}}\mathbf{1} \\ -\zeta_{\text{lb}}\mathbf{1} \\ \zeta_{\text{ub}} \\ -\zeta_{\text{lb}} \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

where for each $\boldsymbol{\ell} \in \mathcal{L}$, $\boldsymbol{f}_{\boldsymbol{\ell}} \in \mathbb{R}^{|\mathcal{K}|}$ is the vector of maximum flows for each scenario in $|\mathcal{K}|$.

C Solving for deterministic strategy

For a deterministic interdiction plan, problem (7) reduces to the following MILP:

$$\begin{array}{lll}
\underset{\substack{\boldsymbol{\ell}, t, \zeta, \boldsymbol{w}^+, \boldsymbol{w}^-, \boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \\ \beta, \boldsymbol{\Delta}, \{\boldsymbol{\lambda}_k, \boldsymbol{\nu}_k\}_{k\in\mathcal{K}}, \boldsymbol{\Upsilon}_k}}{\text{minimize}} & t & \\
\text{subject to} & \zeta + \Gamma\sqrt{|\mathcal{K}|}\beta + \Gamma \sum_{k'\in\mathcal{K}} (w_{k'}^+ + w_{k'}^-) + \dfrac{1}{1-\alpha}\sum_{\boldsymbol{\ell}\in\mathcal{L}} \Delta_{\boldsymbol{\ell},k} & \\
& +(\hat{\boldsymbol{q}} - \boldsymbol{e}_k)^\top(\boldsymbol{w}^+ - \boldsymbol{w}^- + \boldsymbol{\theta}^+ - \boldsymbol{\theta}^-) \le t & \forall k \in \mathcal{K} \\
& \theta_k^+ + \theta_k^- - \beta = 0 & \forall k \in \mathcal{K} \\
& \boldsymbol{w}^+ \ge 0,\, \boldsymbol{w}^- \ge 0,\, \boldsymbol{\theta}^+ \ge 0,\, \boldsymbol{\theta}^- \ge 0 & \\
& \Delta_k \ge \boldsymbol{c}_k^\top \boldsymbol{\lambda}_k - \boldsymbol{c}_k^\top \boldsymbol{\Upsilon}_k - \zeta & \forall k \in \mathcal{K} \\
& \boldsymbol{\lambda}_k + N^\top \boldsymbol{\upsilon}_k - \boldsymbol{d} \ge 0 & \forall k \in \mathcal{K} \\
& 0 \le \boldsymbol{\lambda}_k \le 1 & \forall k \in \mathcal{K} \\
& \boldsymbol{\Upsilon}_k \le \boldsymbol{\ell} & k \in \mathcal{K} \\
& \boldsymbol{\Upsilon}_k \le \boldsymbol{\lambda}_k & \forall k \in \mathcal{K} \\
& \boldsymbol{\Upsilon}_k \ge \boldsymbol{\lambda}_k + \boldsymbol{\ell} - 1 & \forall k \in \mathcal{K} \\
& \boldsymbol{\Upsilon}_k \ge \mathbf{0} & \forall k \in \mathcal{K} \\
& \Delta_k \ge 0 & \forall k \in \mathcal{K} \\
& \boldsymbol{\ell} \in \{0,1\}^{|E|} & \\
& \mathbf{1}^\top \boldsymbol{\ell} \le B, & \\
& 0 \le \zeta \le \bar{\zeta}, &
\end{array}$$

where $\bar{\zeta} := \max_{k\in\mathcal{K}} f_{\mathbf{0},k}$.

The spatial branch and bound algorithm can be applied to any polyhedral distributional ambiguity set because Fenchel robust counterpart of the constraint in (4b) is linear programming representable for polyhedral sets. As a result, McCormick inequalities combined with RLT lead to a linear relaxation of the interdictor's DRMFNI problem which can be solved using our column generation algorithm.

D Constraint generation algorithm

The DRMFNI problem in (4) can be solved by an alternative method based on a constraint generation procedure that exploits the subset-based representation of CVaR given in Künzi-Bay and Mayer (2006) and Fábián (2008). An equivalent subset-based representation of problem (4) is given as follows:

$$\begin{aligned}
\underset{\boldsymbol{u},\zeta,t_1,\boldsymbol{\eta}}{\text{minimize}} \quad & \zeta + \frac{1}{1-\alpha} t_1 && && \text{(28a)}\\
\text{subject to} \quad & \sum_{(\ell,k)\in\mathcal{J}} q_k u_{\boldsymbol{\ell}} f_{\boldsymbol{\ell},k} - q_k \eta_{\boldsymbol{\ell}} \leq t_1 && \forall \mathcal{J} \in \Theta(\mathcal{L}\times\mathcal{K}),\ \forall \boldsymbol{q} \in \mathcal{Q} && \text{(28b)}\\
& t_1 \geq 0 && && \text{(28c)}\\
& \eta_{\boldsymbol{\ell}} = u_{\boldsymbol{\ell}} \zeta && \forall \boldsymbol{\ell} \in \mathcal{L} && \text{(28d)}\\
& \boldsymbol{u} \geq 0 && && \text{(28e)}\\
& \mathbf{1}^\top \boldsymbol{u} = 1 && && \text{(28f)}\\
& \bar{\zeta}_{\text{lb}} \leq \zeta \leq \bar{\zeta}_{\text{ub}}, && && \text{(28g)}
\end{aligned}$$

where $\Theta(\mathcal{L}\times\mathcal{K})$ is the power set of $\{(\boldsymbol{\ell},k)|\ \ell\in\mathcal{L},\ k\in\mathcal{K}\}$.

Since the size of $\Theta(\mathcal{L}\times\mathcal{K})$ is exponential with respect to $|\boldsymbol{\ell}|\cdot|\mathcal{K}|$ and given that we expect at optimum that constraint (28b) will be active only for a few $\mathcal{J}$'s and $\boldsymbol{q}$'s, we can use the constraint generation procedure described in Algorithm 4 to solve problem (28). In words, the algorithm can be described as follows: at any iteration i, we solve the relaxed master problem (29) to obtain the optimal u^*, ζ^*, t_1^* and progressively add constraints that lead to tighter relaxations of problem (28).

$$\begin{aligned}
\text{(MP subset)}: \quad \underset{\boldsymbol{u},\zeta,t_1,\boldsymbol{\eta}}{\text{minimize}} \quad & \zeta + \frac{1}{1-\alpha} t_1 && && \text{(29a)}\\
\text{subject to} \quad & \sum_{(\ell,k)\in\mathcal{J}_j} q_k^j u_{\boldsymbol{\ell}} f_{\boldsymbol{\ell},k} - q_k^j \eta_{\boldsymbol{\ell}} \leq t_1 && \forall j \in \{1,2,\ldots,i-1\} && \text{(29b)}\\
& \text{(28c)} - \text{(28g)} && && \text{(29c)}
\end{aligned}$$

Algorithm 4 Constraint generation algorithm for solving problem (4)

procedure ConstraintGeneration(ϵ)
  $i \leftarrow 1$, $t_1^* = -\infty$ and $t_2^* = \infty$
  while $t_2^* - t_1^* \geq (1-\alpha)\epsilon$ **do**
    Solve problem (29) to get optimal value opt* and optimal $\hat{\boldsymbol{u}}^*$, t_1^*, and ζ^*
    $\mathcal{J}_i \leftarrow \{(\ell,k)|f_{\ell,k} - \zeta^* > 0,\ \hat{u}_\ell^* > 0\}$
    $\boldsymbol{q}^i \leftarrow \text{argmax}_{\boldsymbol{q}\in\mathcal{Q}} \sum_{(\ell,k)\in\mathcal{J}_i} q_k(\hat{u}_{\boldsymbol{\ell}}^* f_{\boldsymbol{\ell},k} - \zeta^*)$
    $t_2^* \leftarrow \sum_{(\ell,k)\in\mathcal{J}_i} q_k^i(\hat{u}_{\boldsymbol{\ell}}^* f_{\boldsymbol{\ell},k} - \zeta^*)$
    $i \leftarrow i+1$
  end while
  return $\hat{\boldsymbol{u}}^*, \zeta^* + \frac{t_2^*}{1-\alpha}$
end procedure

Table 8: Comparison of computation times for the Gurobi's bilinear solver, constraint generation combined with Gurobi's bilinear solver, and our spatial branch and bound algorithm with- and without column generation.

	$\|\mathcal{L}\|$	$\|\mathcal{K}\|$	Gurobi	ConGen+ Gurobi			spatial B&B	spatial B&B +CG
					Last iter.			
			(s)	(s)	(s)	avg. # iter.	(s)	(s)
Nobel-US $\|E\| = 42$, $\|V\| = 14$, $B = 1$	42	100	0.54	4.72	0.21	17.4	1.17	2.57
	42	150	0.85	6.03	0.22	19.6	2.62	6.63
	42	200	1.61	7.18	0.22	19.1	3.40	9.06
Sioux-Falls $\|E\| = 76$, $\|V\| = 24$, $B = 1$	76	100	0.71	2.77	0.19	9.2	0.95	2.42
	76	150	1.25	3.26	0.19	8.3	1.38	3.66
	76	200	2.23	2.96	0.19	9.8	1.72	6.66
Grid Network (5×5) $\|E\| = 50$, $\|V\| = 27$, $B = 2$	496	100	31.54	59.20**	3.48	37.6	35.87	7.16
	496	150	67.64*	96.90**	3.28	59.0	31.16	10.47
	496	200	73.70	67.85**	3.37	49.6	43.32	13.25

* average computed on nine instances as one failed to converge in 10 min
** average computed on eighth instances as two failed to converge in 10 min

E Realistic network instances from the literature

The topologies of the network instances studied in LeBlanc et al. (1975) and Orlowski et al. (2010) are presented respectively in figures 4 and 5.

F Sensitivity of computation time to the choice of a reference distribution

We randomly generated 10 instances of size $15 \times 15 \times 15$ for the grid network given in Figure 1 for each level of $\Gamma \in \{0.5|\mathcal{K}|, 0.1|\mathcal{K}|, 0.15|\mathcal{K}|\}$ and the scenarios for the capacities of the arcs are generated from the factor model described in Section 5. Instead of assuming that the reference distribution in our DRMNFNI problem is the empirical distribution of the sample of scenarios, we randomly generate the reference distribution of each problem instance using a Dirichlet distribution with parameters $\beta_i := \beta \in \{0.1, 0.5, 1, \infty\}$, where β controls the concentration Dirchlet density around $\hat{\boldsymbol{q}} = (1/|\mathcal{K}|)\mathbf{1}$. For a budget of interdiction $B = 4$ and risk-aversion parameter $\alpha = 0.05$, the average and standard deviation of computation time for different levels of Γ and β are reported in Table 9. We can see that the statistics of computation time are not sensitive to the variability (controlled by β) of the reference distribution.

References

Ahuja RK, Magnanti TL, Orlin JB (1993) *Network Flows: Theory, Algorithms, and Applications* (Upper Saddle River, NJ, USA: Prentice-Hall, Inc.).

Al-Khayyal FA, Falk JE (1983) Jointly constrained biconvex programming. *Mathematics of Operations Research* 8(2):273–286.

Assimakopoulos N (1987) A network interdiction model for hospital infection control. *Computers in Biology and Medicine* 17(6):413–422.

Atamtürk A, Deck C, Jeon H (2020) Successive quadratic upper-bounding for discrete mean-risk minimization and network interdiction. *INFORMS Journal on Computing* 32(2):346–355.

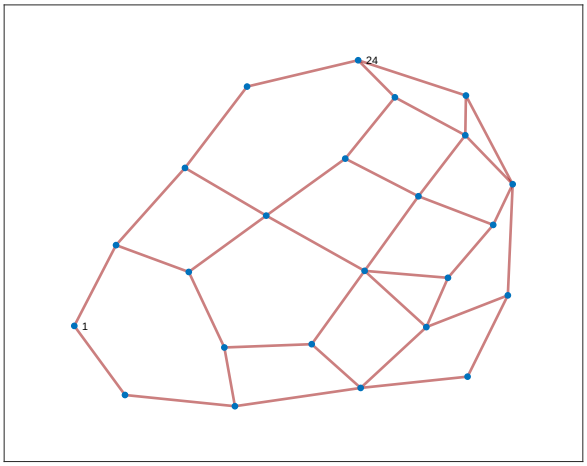

Figure 4: Sioux-Falls network from LeBlanc et al. (1975).

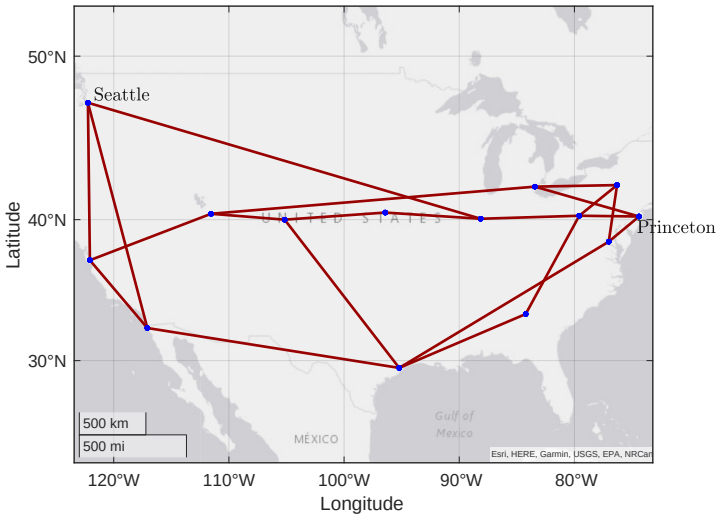

Figure 5: Nobel-US network from Orlowski et al. (2010).

Table 9: Statistics (in the format of "Average (Standard deviation)") of computation times of spatial branch and bound algorithm for different reference distributions and different levels of uncertainty (Γ) for 10 randomly generated instances.

Concentration parameter (β)	Level of Uncertainty (Γ) (in % of $\lvert\mathcal{K}\rvert$)		
	5% (s)	10% (s)	15% (s)
0.1	9.14 (24.79)	11.57 (27.66)	12.55 (30.37)
0.5	9.89 (26.41)	12.59 (30.30)	12.70 (30.61)
1	9.68 (25.78)	12.75 (30.73)	12.86 (31.17)
∞	9.80 (26.07)	12.67 (30.67)	12.77 (30.97)

Bayraksan G, Love DK (2015) Data-driven stochastic programming using phi-divergences. *The Operations Research Revolution, INFORMS TutORials in Operations Research* 1–19.

Ben-Tal A, den Hertog D, De Waegenaere A, Melenberg B, Rennen G (2013) Robust solutions of optimization problems affected by uncertain probabilities. *Management Science* 59(2):341–357.

Ben-Tal A, den Hertog D, Vial JP (2015) Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming* 149(1):265–299.

Ben-Tal A, Nemirovski A (2000) Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming* 88(3):411–424.

Ben-Tal A, Nemirovski A (2008) Selected topics in robust convex optimization. *Mathematical Programming* 112(1):125–158.

Bertsimas D, Nasrabadi E, Orlin JB (2016) On the power of randomization in network interdiction. *Operations Research Letters* 44(1):114–120.

Borrero JS, Prokopyev OA, Sauré D (2019) Sequential interdiction with incomplete information and learning. *Operations Research* 67(1):72–89.

Chandraker M, Kriegman D (2008) Globally optimal bilinear programming for computer vision applications. *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 1–8 (Anchorage, AK: IEEE Computer Society).

Chicoisne R, Ordóñez F, Espinoza D (2018) Risk averse shortest paths: A computational study. *INFORMS Journal on Computing* 30(3):539–553.

Cormican KJ, Morton DP, Wood RK (1998) Stochastic network interdiction. *Operations Research* 46(2):184–197.

Delage E, Kuhn D, Wiesemann W (2019) "Dice"-sion–making under uncertainty: When can a random decision reduce risk? *Management Science* 65(7):3282–3301.

Delage E, Saif A (2021) The value of randomized solutions in mixed-integer distributionally robust optimization problems. *INFORMS Journal on Computing* URL `https://doi.org/10.1287/ijoc.2020.1042`, ePub ahead of print, May 13.

Desrosiers J, Lübbecke ME (2005) A primer in column generation. Desaulniers G, Desrosiers J, Solomon MM, eds., *Column Generation*, 1–32 (Boston, MA: Springer US).

Duchi JC, Glynn PW, Namkoong H (2021) Statistics of robust optimization: A generalized empirical likelihood approach. *Mathematics of Operations Research* 46(3):946–969.

Ellsberg D (1961) Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics* 75(4):643–669.

Epstein LG (1999) A definition of uncertainty aversion. *The Review of Economic Studies* 66(3):579–608.

Fábián CI (2008) Handling cvar objectives and constraints in two-stage stochastic models. *European Journal of Operational Research* 191(3):888–911.

Hajinezhad D, Hong M (2019) Perturbed proximal primal–dual algorithm for nonconvex nonsmooth optimization. *Mathematical Programming* 176(1):207–245.

Hajinezhad D, Shi Q (2018) Alternating direction method of multipliers for a class of nonconvex bilinear optimization: convergence analysis and applications. *Journal of Global Optimization* 70(1):261–288.

Holzmann T, Smith JC (2021) The shortest path interdiction problem with randomized interdiction strategies: Complexity and algorithms. *Operations Research* 69(1):82–99.

Israeli E, Wood RK (2002) Shortest-path network interdiction. *Networks* 40:97–111.

Jain M, Tsai J, Pita J, Kiekintveld C, Rathi S, Tambe M, Ordóñez F (2010) Software assistants for randomized patrol planning for the lax airport police and the federal air marshal service. *INFORMS Journal on Applied Analytics* 40(4):267–290.

Janjarassuk U, Linderoth J (2008) Reformulation and sampling to solve a stochastic network interdiction problem. *Networks* 52(3):120–132.

Ji R, Lejeune MA (2021) Data-driven optimization of reward-risk ratio measures. *INFORMS Journal on Computing* 33(3):1120–1137.

Künzi-Bay A, Mayer J (2006) Computational aspects of minimizing conditional value-at-risk. *Computational Management Science* 3(1):3–27.

Lam H (2019) Recovering best statistical guarantees via the empirical divergence-based distributionally robust optimization. *Operations Research* 67(4):1090–1105.

LeBlanc LJ, Morlok EK, Pierskalla WP (1975) An efficient approach to solving the road network equilibrium traffic assignment problem. *Transportation Research* 9(5):309–318.

Lei X, Shen S, Song Y (2018) Stochastic maximum flow interdiction problems under heterogeneous risk preferences. *Computers & Operations Research* 90:97–109.

Liberti L, Pantelides CC (2006) An exact reformulation algorithm for large nonconvex nlps involving bilinear terms. *Journal of Global Optimization* 36(2):161–189.

Loizou N (2015) Distributionally robust game theory, M.Sc thesis, Imperial College London.

Magliocca NR, McSweeney K, Sesnie SE, Tellman E, Devine JA, Nielsen EA, Pearson Z, Wrathall DJ (2019) Modeling cocaine traffickers and counterdrug interdiction forces as a complex adaptive system. *Proceedings of the National Academy of Sciences* 116(16):7784–7792.

McCormick GP (1976) Computability of global solutions to factorable nonconvex programs: Part I–Convex underestimating problems. *Mathematical Programming* 10(1):147–175.

McLay LA, Lloyd JD, Niman E (2011) Interdicting nuclear material on cargo containers using knapsack problem models. *Annals of Operations Research* 187(1):185–205.

Mohajerin Esfahani P, Kuhn D (2018) Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming* 171(1):115–166.

Orlowski S, Wessäly R, Pióro M, Tomaszewski A (2010) Sndlib 1.0—survivable network design library. *Networks: An International Journal* 55(3):276–286, URL `http://sndlib.zib.de`.

Pay BS, Merrick JR, Song Y (2019) Stochastic network interdiction with incomplete preference. *Networks* 73(1):3–22.

Postek K, den Hertog D, Melenberg B (2016) Computationally tractable counterparts of distributionally robust constraints on risk measures. *SIAM Review* 58(4):603–650.

Rockafellar R, Uryasev S (2000) Optimization of conditional value-at-risk. *Journal of Risk* 2:21–41.

Royset JO, Wood RK (2007) Solving the bi-objective maximum-flow network-interdiction problem. *INFORMS Journal on Computing* 19(2):175–184.

Sherali HD, Alameddine A (1992) A new reformulation-linearization technique for bilinear programming problems. *Journal of Global Optimization* 2(4):379–410.

Sion M (1958) On general minimimax theorems. *Pacific Journal of Mathematics* 8(1):171–176.

Smith JC, Prince M, Geunes J (2013) Modern network interdiction problems and algorithms. Pardalos PM, Du DZ, Graham RL, eds., *Handbook of Combinatorial Optimization*, 1949—-1987 (New York, NY: Springer New York).

Smith JC, Song Y (2020) A survey of network interdiction models and algorithms. *European Journal of Operational Research* 283(3):797–811.

Smith JE, Winkler RL (2006) The optimizer's curse: Skepticism and postdecision surprise in decision analysis. *Management Science* 52(3):311–322.

Song Y, Shen S (2016) Risk-averse shortest path interdiction. *INFORMS Journal on Computing* 28(3):527–539.

Wood RK (1993) Deterministic network interdiction. *Mathematical and Computer Modelling* 17(2):1–18.

Yang J, Borrero JS, Prokopyev OA, Sauré D (2021) Sequential shortest path interdiction with incomplete information and limited feedback. *Decision Analysis* 18(3):218–244.